THE MODAL LOGICS OF KRIPKE-FEFERMAN TRUTH

CARLO NICOLAI AND JOHANNES STERN

Abstract. We determine the modal logic of fixed-point models of truth and their axiomatizations by Solomon Feferman via Solovay-style completeness results. Given a fixed-point model $\mathcal{M}$, or an axiomatization S thereof, we find a modal logic M such that a modal sentence φ is a theorem of M if and only if the sentence φ^* obtained by translating the modal operator with the truth predicate is true in $\mathcal{M}$ or a theorem of S under all such translations. To this end, we introduce a novel version of possible worlds semantics featuring both classical and nonclassical worlds and establish the completeness of a family of non-congruent modal logics whose internal logic is nonclassical with respect to this semantics.

1. Introduction

In the aftermath of Gödel's incompleteness theorems, researchers became interested in general properties of the formalized provability predicate. Bernays distilled three such properties known as Hilbert-Bernays derivability conditions but it was Löb [Loe55] who, elaborating on Bernays' work, proposed the three derivability conditions that are commonly thought to aptly characterize the properties of a "natural" provability predicate of recursively enumerable systems extending a sufficiently strong arithmetical theory.[1] The striking feature of Löb's derivability conditions is that they are essentially principles of propositional modal logic. That is, if the provability predicate is replaced by a modal operator, the three derivability conditions can be studied as axioms of systems of modal logic. It is then natural to ask which modal system is the exact modal logic of the provability predicate of recursively enumerable systems extending $\mathrm{I}\Delta_0 + \mathrm{Exp}$ – a question that was answered by Solovay's [Sol76] seminal arithmetical completeness result that established the logic to be the system GL.

Interestingly, in his paper Solovay observes that by considering alternative modal principles and systems we can study the modal logic of other sentential predicates. It is in this spirit that we investigate the modal logic of the truth predicate of fixed-point models of truth in the sense of Kripke [Kri75] and their axiomatizations by Feferman [Fef91, Fef08].[2] To this effect we propose Solovay-style completeness results for a family of truth theories. Let Σ be such a truth theory. Then we determine the modal logic S such that for all $\varphi \in \mathcal{L}_\Box$:

$$S \vdash \varphi \Leftrightarrow \text{for all realizations } * \, (\Sigma \vdash \mathfrak{I}^*(\varphi)).$$

Here $*$ is a function that maps the propositional variable of the modal language to sentences of the language of the truth theory; $\mathfrak{I}$ is the truth-interpretation, that is translation that commutes with the logical connectives and translates the modal

2000 *Mathematics Subject Classification.* Primary 03Bxx; Secondary 03B45; 03B50.

[1] $\mathrm{I}\Delta_0 + \mathrm{Exp}$ can be considered to be a safe lower bound. Here Exp stands for the sentence asserting the totality of the exponential function $x \mapsto 2^x$. See [HP98] for details.

[2] For a general overview of axiomatizations of fixed-point semantics, see [Hal14, §15-17].

operator $\Box$ by the truth predicate:

$$\mathfrak{I}^*(\Box\varphi) = \mathrm{T}\ulcorner\mathfrak{I}^*(\varphi)\urcorner.$$

Some initial research in this direction has already been undertaken by Czarnecki and Zdanowski [CZ19] and Standefer [Sta15] who study theories of truth inspired by revision theoretic approaches [GB93].[3] Czarnecki and Zdanowski, and Standefer determine, albeit in slightly different guise, the modal logic of nearly stable truth, that is, the modal logic of the axiomatic truth theory Friedman-Sheard [FS87, Hal94], to be the modal logic $\mathsf{KDD_c}$.

Determining the precise modal logics of the truth predicate in a truth theory is interesting for several respects.

Modal Analysis of Self-Reference. Solovay's result and provability logic more generally improved our understanding of the notion of provability and provided a modal analysis of self-reference. Analogously, determining the modal logic of the truth predicate improves our understanding of the notion of truth employed by the truth theory at stake. Of course, modal logics of truth do not lend themselves to the modal analysis of self-reference in the way provability logics do: in particular, a De Jongh/Sambin-style fixed-point theorem will not be provabl for the modal logics of truth.[4]

However, the study of modal logics of truth contributes to the modal analysis of self-reference in different ways. First, truth-theoretic completeness results yield modal logics which can be consistently enriched, syntactically and semantically, by fixed points for modal formulas in the sense of Smorynski's Diagonalization Operator Logic [Smo85]. Secondly, we show that some of these logics are maximal in this sense: there will be no stronger modal logics that can be consistently augmented by such fixed points.

Maximally Consistent Sets of Truth Principles. The previous feature leads to a second point of interest in determining the modal logics of the truth predicate. An important question in the logical and philosophical study of truth is to determine principled ways of selecting maximally consistent approximations to the full T-schema: '*A* is true' if and only if *A*. In their important study, Friedman and Sheard [FS87] devised a number of basic approximations to the T-schema and determined maximally consistent sets of principles relative to such approximations. However, Friedman and Sheard's results were relative to the specific sets of basic approximations to the T-schema they started with: such results leave open, for instance, whether one could extend their starting sets with different principles to obtain consistent extensions of their maximally consistent sets. In contrast, as a direct consequence of our results in Section 5.2, we obtain a number of absolute results, that is, we show that the modal logics of the truth theories $\mathsf{KF}+\mathsf{CN}$, $\mathsf{KF}+\mathsf{CM}$, WKFC, and DT are maximal.[5] This implies that the truth principles of these theories form a maximally consistent set in the absolute sense. It will not be possible to strengthen the modal logic of these theories by any truth principle.

Besides addressing interesting conceptual questions, the project also displays technical challenges.

Diagonalization and T*-sentences.* It may seem that in the presence of suitable axiomatic theories of truth determining the corresponding modal logic is a simple matter: we obtain the modal logic by replacing the truth predicate by the modal

[3]For related work see also Stern [Ste15] who connects various diagonal modal logics [Smo85] to truth theories via the truth interpretation.

[4]This follows from the fact that all modal logics of truth will be sublogics of the so-called identity logic, that is, the modal logic that interprets the modal operator as an identity operator. For sake of consistency the identity logic cannot have the fixed-point property.

[5]Similarly, [CZ19] show the modal logic $\mathsf{KDD_c}$ to be maximal.

operator and the individual variable ranging over sentences by propositional variables of the modal operator language in the axioms of the theory. This picture is incorrect for at least two reasons.

First, the truth theory will be formulated over a theory of syntax, usually some arithmetical theory such as $\mathrm{I}\Delta_0 + \mathrm{Exp}$ or extensions thereof, which will equip the truth theory with further expressive strength.[6] For instance, in the truth theory we can prove the diagonal lemma, which is usually not possible within a modal logic. But by using the diagonal lemma we may prove further modal principles, which we cannot prove in the modal logic extracted in the simplistic way from our theory. Indeed, the formalized Löb's theorem provides a neat example to this effect: if we have the diagonal lemma, it can be proved on the basis of Löb's three derivability conditions, but we cannot prove the modal Löb principle in the modal logic $\mathsf{K4}$.

Second, in truth theories one usually stipulates the T-sentences for atomic sentences of the language without the truth predicate. But if we consider the parallel principle

$$(\mathsf{TB}) \qquad \Box p \leftrightarrow p$$

in the context of a modal operator logic we run into trouble: modal logics are closed under the rule of uniform substitution, which means that (TB) would hold for all sentences φ of the language with the truth predicate. But then, as a consequence of the Liar paradox, no interesting classical truth theory can have a modal logic that assumes (TB).[7] Nonetheless, already the T-sentences for atomic sentences in the language without the truth predicate – which are required in providing non-trivial axiomatizations of the truth predicate – have an impact on the modal logic of the theory and the exact impact cannot be immediately read off the axioms of the truth theory. In fact, we will show that such T-sentences force the modal logic of theories of Kripke-Feferman truth to comprise the axiom

$$(\textsc{faith}^{\Box}) \qquad \Box\varphi \land \neg\Box\neg\varphi \to \varphi,$$

which does not appear in the usual list of axioms of these truth theories. Thus $\textsc{faith}^{\Box}$, for slightly different reasons, plays a similar role as Löb's principle in GL: it is easily derivable in our theories of truth, but it needs to be added to their modal logics as an additional axiom.

Novel possible worlds Structures. As mentioned this paper is concerned with so-called Kripkean fixed-point theories of truth and their axiomatizations. The peculiar feature of these theories is that the truth predicate is in some respect nonclassical: depending on the particular version of the theory under consideration sentences may be neither true nor false, or both true and false. As a consequence, the so-called internal logic of the truth predicate, that is the logic holding within the scope of the truth predicate, will be nonclassical, which, in turn, forces the modal logic to be non-normal and, indeed, non-congruent.[8] This has the further consequence that we cannot appeal to standard modal semantics for investigating the modal systems under consideration but have to introduce a novel version of

[6]We highlight, however, that our results will also apply to base theories weaker than $\mathrm{I}\Delta_0+\mathrm{Exp}$, such as Buss's S^1_2 or $\mathsf{I}\Delta_0 + \Omega_1$.

[7]It is precisely in this sense that our modal logics diverge from Feferman's [Fef84] 'type-free modal theory': while Feferman also picks up on the modal character of the various truth principles and formulates them using a modal operator, he does this over a specific theory and adopts TB, that is, Feferman is to be taken literally when he calls his system a 'modal theory' rather than a 'modal logic'. His 'modal theory' is not closed under uniform substitution and should arguably not be called a logic.

[8]Non-congruent modal logics are the nonclassical modal logics in the sense of [Seg71] and [Che80].

possible world semantics, akin to impossible world semantics [Pri08], in which we allow for both classical and nonclassical worlds.[9]

1.1. **Plan of the paper.** In §2 we introduce nonclassical modal logics and their semantics, which extend the main (fully structural) nonclassical logics considered in the literature on theories of truth. These include the four-valued logic of first-degree entailment and its three-valued paraconsistent and paracomplete extensions. As we will show later in the paper, the nonclassical modal logic are the modal logics of the internal theory or logic of the systems of Kripke-Feferman truth. §3 presents the classical modal logics that will be shown to be exactly the modal logics of Kripke-Feferman theories. Since these logics are classical but have a nonclassical internal logics, providing a semantics for them is a nontrivial task. To this end, we introduce special frames in which classical worlds 'see' a unique nonclassical world, and prove the completeness of our modal logics with respect to this semantics. In §4, we move on to introduce Kripkean fixed point semantics and the systems of Kripke-Feferman truth. §5 contains the main results of the paper. We first establish the modal logic of the basic system KF and its extensions with completeness of consistency axioms for the truth predicate. Then we consider – by providing a different realization – a stronger form of truth theoretic completeness that holds for KF + CN, KF + CM, WKFC, DT, and all of their consistent extensions. We conclude the paper by pointing to some further research.

2. NONCLASSICAL MODAL LOGICS FOR TRUTH

Our basic language is a standard modal operator language $\mathcal{L}_\Box$, which is built over a countable set Prop of propositional variables $p_0, p_1, p_2, \ldots$:

$$\varphi ::= p_j \mid \bot \mid \top \mid \neg\varphi \mid \Box\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi$$

with $p_j \in \text{Prop}$ for $j \in \omega$. The language $\mathcal{L}_0$ is the language obtained by removing $\Box$ from $\mathcal{L}_\Box$. $\mathcal{L}_\Box^{\twoheadrightarrow}$ is obtained by adding to $\mathcal{L}_\Box$ a binary connective $\twoheadrightarrow$. In the context of our nonclassical modal logics, $\Diamond$ can be defined as usual as $\neg\Box\neg$, but this *will not* be the case in the classical modal logics investigated in the following sections. A *literal* is defined as a propositional variable or a negated propositional variable.

We will work with several nonclassical logics that support naïve truth and that amount to the internal logics of the classical systems of truth we are ultimately interested in. We formulate our systems in a sequent calculus, where $\Gamma, \Delta, \Theta, \Lambda \ldots$ are *finite sets* of formulas of $\mathcal{L}_0$. We denote with $\text{Prop}(\varphi)$ the set of propositional variables of φ (resp. $\text{Prop}(\Gamma)$ for the set of propositional variables in the set of formulas Γ). For X a set of sentences of $\mathcal{L}_\Box$, we let $\Box X := \{\Box\varphi \mid \varphi \in X\}$.

We work with sequent calculus formulations for first-degree entailment (FDE), symmetric Kleene logic (KS3), Strong Kleene (K3), the logic of paradox (LP), Weak Kleene (B3), and Feferman-Aczel logic (F3), which is a slight modification of B3 with a primitive conditional [AB75, Kle52, Cos74, CC13, Acz80, Fef08]. Details of the systems are given in Appendix A.

We now introduce some modal extensions of our systems.[10]

[9]Of course, devising a new possible worlds framework is not the only challenge. To obtain the truth-theoretic completeness of our modal logics, such possible worlds structures need to be encoded in suitable fixed-point models.

[10]Modal logics based on K3 are studied in [JT96]. Their framework forms the basis of the logical systems of nonclassical modal logic considered in this paper. An FDE-based version of K is extensively studied in [OW10] and labelled BK. Also, consistent and complete extensions of BK have been investigated by [OS20] and [OS16].

Definition 1. *For $S \in \{\mathsf{K3}, \mathsf{B3}, \mathsf{LP}, \mathsf{KS3}, \mathsf{FDE}, \mathsf{F3}\}$, the systems $S_\Box$ in the language $\mathcal{L}_\Box$ ($\mathcal{L}_\Box^\rightarrow$) are defined by adding to S the rules:*

$$(\Box\text{L})\ \frac{\varphi, \Gamma \Rightarrow \Delta}{\Box\varphi, \Gamma \Rightarrow \Delta} \qquad (\Box\text{R})\ \frac{\Gamma \Rightarrow \Delta, \varphi}{\Gamma \Rightarrow \Delta, \Box\varphi}$$

$$(\neg\Box\text{L})\ \frac{\neg\varphi, \Gamma \Rightarrow \Delta}{\neg\Box\varphi, \Gamma \Rightarrow \Delta} \qquad (\neg\Box\text{R})\ \frac{\Gamma \Rightarrow \Delta, \neg\varphi}{\Gamma \Rightarrow \Delta, \neg\Box\varphi}$$

The adequacy of the logics $S_\Box$ with respect to the possible worlds semantics introduced below follows from a more general result concerning a nonclassical modal logic, which can be thought of as the modal analogue of the classical propositional logic K.

Definition 2. *For $S \in \{\mathsf{K3}, \mathsf{B3}, \mathsf{LP}, \mathsf{KS3}, \mathsf{FDE}, \mathsf{F3}\}$, the systems $S_\blacksquare$ in $\mathcal{L}_\Box$ ($\mathcal{L}_\Box^\rightarrow$) are obtained by replacing (*REF*) with the initial sequent $\Gamma, \varphi \Rightarrow \varphi, \Delta$ for all $\varphi \in \mathcal{L}_\Box$ ($\mathcal{L}_\Box^\rightarrow$), and by adding to S the rules:*

$$(\blacksquare\text{L})\ \frac{\Gamma, \neg\varphi \Rightarrow \neg\Delta}{\Box\Gamma, \neg\Box\varphi \Rightarrow \neg\Box\Delta} \qquad (\blacksquare\text{R})\ \frac{\Gamma \Rightarrow \varphi, \neg\Delta}{\Box\Gamma \Rightarrow \Box\varphi, \neg\Box\Delta}$$

We apply the following notational conventions for T one of our modal systems:

- Derivations in T are at most binary branching finite trees labelled with sequents. Leaves are axioms – the relevant instance of reflexivity, ($\bot$), ($\top$), and the remaining nodes are obtained by applications of the rules of inference of T. For T one of the logics above, $T \vdash \Gamma \Rightarrow \Delta$ stands for the existence of a derivation whose root is labelled by $\Gamma \Rightarrow \Delta$.
- We can extend the above notion of derivability to *arbitrary* sets of formulas: for Γ, Δ arbitrary sets of formulas, we write $T \vdash \Gamma \Rightarrow \Delta$ iff there are finite $\Gamma_0 \subseteq \Gamma$ and $\Delta_0 \subseteq \Delta$ such that $T \vdash \Gamma_0 \Rightarrow \Delta_0$.
- The *length* of a derivation can be defined as the number of nodes in the maximal branch of the derivation tree minus one. We write $T \vdash^n \Gamma \Rightarrow \Delta$ if the length of the derivation of $\Gamma \Rightarrow \Delta$ in T is $\leq n$.

By our definition of sequent, contraction is trivially admissible in our logics. In addition, by straightforward induction on the length of the derivations in the appropriate systems, we have:

Lemma 3 (Reflexivity, Weakening).

(i) *For $S \in \{\mathsf{K3}, \mathsf{LP}, \mathsf{KS3}, \mathsf{FDE}, \mathsf{B3}\}$ as in the previous definition and $\Gamma, \Delta \subseteq \mathcal{L}_\Box$,*
  (a) For all $\varphi \in \mathcal{L}_\Box$, $S_\Box \vdash \Gamma, \varphi \Rightarrow \varphi, \Delta$
  (b) If $S_\Box \vdash^n \Gamma \Rightarrow \Delta$, then $S_\Box \vdash^n \Gamma, \Gamma_0 \Rightarrow \Delta, \Delta_0$ for Γ_0, Δ_0 finite.
  (c) if $S_\blacksquare \vdash^n \Gamma \Rightarrow \Delta$, then $S_\blacksquare \vdash^n \Gamma, \Gamma_0 \Rightarrow \Delta, \Delta_0$ for Γ_0, Δ_0 finite.

(ii) *For $\varphi \in \mathcal{L}_\Box^\rightarrow$, and $\Gamma, \Delta \subseteq \mathcal{L}_\Box^\rightarrow$:*
  (a) $\mathsf{F3}_\Box \vdash \Gamma, \varphi \Rightarrow \varphi, \Delta$
  (b) If $\mathsf{F3}_\Box \vdash^n \Gamma \Rightarrow \Delta$, then $\mathsf{F3}_\Box \vdash^n \Gamma, \Gamma_0 \Rightarrow \Delta, \Delta_0$.
  (c) if $\mathsf{F3}_\blacksquare \vdash^n \Gamma \Rightarrow \Delta$, then $\mathsf{F3}_\blacksquare \vdash^n \Gamma, \Gamma_0 \Rightarrow \Delta, \Delta_0$.

2.1. **Semantics.** Next we introduce a possible worlds semantics for the propositional systems just defined. The main difference with standard possible models lies in the use of nonclassical valuation functions, which then give rise to nonclassical semantic clauses of the connectives and consequence relations [JT96, Pri08].

Definition 4.

(i) *A* frame *is a pair* (Z, R) *where* Z *is a nonempty set and* R *is a binary relation on* Z*.*

(ii) *A* four-valued *valuation for* $\mathcal{L}_\Box$ *is a function* $V\colon \mathrm{Prop} \times Z \to \{0, 1, \mathsf{n}, \mathsf{b}\}$*, where* $\{0, 1, \mathsf{n}, \mathsf{b}\}$ *is the set of truth values:*
- *a* consistent *valuation is a function* $V\colon \mathrm{Prop} \times Z \to \{0, 1, \mathsf{n}\}$*;*
- *a* complete *valuation is a function* $V\colon \mathrm{Prop} \times Z \to \{0, 1, \mathsf{b}\}$*;*
- *a* symmetric *valuation assigns, at every* $z \in Z$*, values in exactly one of* $\{1, 0, \mathsf{n}\}$ *or* $\{1, 0, \mathsf{b}\}$.

(iii) *A* model $\mathcal{M}$ *is a triple* (Z, R, V)*, with* (Z, R) *a frame and* V *a valuation. A model so-defined is* based on (Z, R)*.*
- *A model* $\mathcal{M} = (Z, R, V)$ *is* consistent *if* V *is consistent.*
- *A model* $\mathcal{M} = (Z, R, V)$ *is* complete *if* V *is complete.*

Let $\preceq$ be the ordering of the truth values $\{0, \mathsf{n}, \mathsf{b}, 1\}$ displayed in the lattice

$$\begin{tikzcd} & 1 & \\ \mathsf{n} \arrow[ur] & & \mathsf{b} \arrow[ul] \\ & 0 \arrow[ul] \arrow[ur] & \end{tikzcd}$$

Moreover, let $\preccurlyeq$ be the ordering $\mathsf{n} \preccurlyeq 0 \preccurlyeq 1$.

Definition 5 (Truth).

(i) *Given a model* $\mathcal{M} = (Z, R, V)$*, and* $s \in \{\mathsf{fde}, \mathsf{k3}, \mathsf{lp}, \mathsf{ks3}\}$*, an* s-interpretation *extends the valuation* V *by assigning to each sentence of* $\mathcal{L}_\Box$ *a truth value:*

$$|p|_s^{\mathcal{M},z} = V_z(p) \qquad |\neg\varphi|_s^{\mathcal{M},z} = \begin{cases} 0 & \textit{if } |\varphi|_s^{\mathcal{M},z} = 1 \\ 1 & \textit{if } |\varphi|_s^{\mathcal{M},z} = 0 \\ |\varphi|_s^{\mathcal{M},z} & \textit{otherwise.} \end{cases}$$

$$|\varphi \wedge \psi|_s^{\mathcal{M},z} = \inf_\preceq\{|\varphi|_s^{\mathcal{M},z}, |\psi|_s^{\mathcal{M},z}\} \qquad |\varphi \vee \psi|_s^{\mathcal{M},z} = \sup_\preceq\{|\varphi|_s^{\mathcal{M},z}, |\psi|_s^{\mathcal{M},z}\}$$

$$|\Box\varphi|_s^{\mathcal{M},z} = \inf_\preceq\{|\varphi|_s^{\mathcal{M},z_0} \mid Rzz_0\}$$

If V *is four-valued, then* $s = \mathsf{fde}$*; if* V *is consistent, then* $s = \mathsf{k3}$*; if* V *is complete, then* $s = \mathsf{lp}$*; finally, if* V *is symmetric, then* $s = \mathsf{ks3}$.

(ii) *Given a consistent* $\mathcal{M} = (Z, R, V)$*, a* $\mathsf{B3}$-interpretation *is given by:*

$$|p|_{\mathsf{b3}}^{\mathcal{M},z} = V_z(p) \qquad |\neg\varphi|_{\mathsf{b3}}^{\mathcal{M},z} = \begin{cases} 0 & \textit{if } |\varphi|_{\mathsf{b3}}^{\mathcal{M},z} = 1 \\ 1 & \textit{if } |\varphi|_{\mathsf{b3}}^{\mathcal{M},z} = 0 \\ |\varphi|_{\mathsf{b3}}^{\mathcal{M},z} & \textit{otherwise.} \end{cases}$$

$$|\varphi \wedge \psi|_{\mathsf{b3}}^{\mathcal{M},z} = \inf_\preccurlyeq(|\varphi|_{\mathsf{b3}}^{\mathcal{M},z}, |\psi|_{\mathsf{b3}}^{\mathcal{M},z}) \qquad |\varphi \vee \psi|_{\mathsf{b3}}^{\mathcal{M},z} = |\neg(\neg\varphi \wedge \neg\psi)|_{\mathsf{b3}}^{\mathcal{M},z}$$

$$|\Box\varphi|_{\mathsf{b3}}^{\mathcal{M},z} = \inf_\preccurlyeq\{|\varphi|_{\mathsf{b3}}^{\mathcal{M},z_0} \mid Rzz_0\}$$

Notice the use of the ordering $\preccurlyeq$ *in this clause.*

(iii) *Again given a consistent* $\mathcal{M} = (Z, R, V)$*, an* $\mathsf{F3}$-interpretation *extends a* $\mathsf{B3}$*-interpretation with the clause:*

$$|\varphi \twoheadrightarrow \psi|_{\mathsf{f3}}^{\mathcal{M},z} = \begin{cases} 1, & \textit{if } |\varphi|_{\mathsf{f3}}^{\mathcal{M},z} = 0 \textit{ or } \inf_\preccurlyeq(|\varphi|_{\mathsf{f3}}^{\mathcal{M},z}, |\psi|_{\mathsf{f3}}^{\mathcal{M},z}) = 1 \\ 0, & \textit{if } |\varphi|_{\mathsf{f3}}^{\mathcal{M},z} = 1 \textit{ and } |\psi|_{\mathsf{f3}}^{\mathcal{M},z} = 0 \\ \mathsf{n}, & \textit{otherwise} \end{cases}$$

Notation.

- Given $\mathcal{M} = (Z, R, V)$ and $s \in \{\mathsf{fde}, \mathsf{lp}, \mathsf{k3}, \mathsf{ks3}, \mathsf{b3}, \mathsf{f3}\}$, we write $\mathcal{M}, z \Vdash_s \varphi$ whenever $|\varphi|_s^{\mathcal{M},z} \in \{1, \mathsf{b}\}$.

In what follows, we will focus on the so-called *local logical consequence relation* in our semantics [BdRV01].

Definition 6 (Consequence). *Let $F = (Z, R)$ be an arbitrary frame, and $\mathfrak{F}$ a class of models based on F. For Γ, Δ sets of sentences of $\mathcal{L}_{\Box}$ and $s \in \{\mathsf{fde}, \mathsf{lp}, \mathsf{k3}, \mathsf{ks3}, \mathsf{b3}, \mathsf{f3}\}$, we have*

$$\Gamma \vDash_s^{\mathfrak{F}} \Delta \text{ iff for all } \mathcal{M} \in \mathfrak{F} \text{ and } z \in Z\text{: if } \mathcal{M}, z \Vdash_s \gamma \text{ for all } \gamma \in \Gamma, \text{ then } \mathcal{M}, z \Vdash_s \delta \text{ for some } \delta \in \Delta.$$

Remark 7. The notion of consequence for FDE and KS3 can be formulated with an extra clause for the anti-preservation of falsity. This is not the case for the stronger logics.

The logics $S_{\blacksquare}$ are adequate with respect to the Kripke semantics introduced above. To prove this we generalize to different evaluation schemata the main strategy applied by [JT96] to modal logics extending $\mathsf{K3}_{\blacksquare}$. In such strategies, the notion of maximally consistent set is replaced with the one of *saturated set.* A saturated set is, roughly, a non-trivial set of formulas closed under the particular logic whose completeness is at stake.[11] The detailed proof of the next claim is provided in Appendix B.

Proposition 8 (Adequacy). *Let $F = (Z, R)$ be an arbitrary frame and S as in Definition 2. Then for any $\Gamma, \Delta \subseteq \mathcal{L}_{\Box}$:*

$$\Gamma \vDash_s^{\mathfrak{F}} \Delta \text{ iff } S_{\blacksquare} \vdash \Gamma \Rightarrow \Delta$$

holds when

- *$\mathfrak{F}$ is the class of four-valued models based on F and S is* FDE*;*
- *$\mathfrak{F}$ is the class of consistent models based on F and S is* K3, B3*, or* F3*;*
- *$\mathfrak{F}$ is the class of complete models based on F and S is* LP*;*
- *$\mathfrak{F}$ is the class of symmetric models based on F and S is* KS3*.*

Notice that, by changing the parameter S in the claims above, we are simultaneously changing both the logic on the right-hand side of the equivalence, and *the evaluation scheme on the left hand side.*

The logics $S_{\blacksquare}$ are, in a sense, the equivalent of the modal logic K in the non-classical settings. Turning to the modal logics $S_{\Box}$ we observe that these logics are precisely the logics of so-called idiosyncratic frames (Figure 2.1).

Definition 9. *Let $F = (Z, R)$ be a frame. F is called* idiosyncratic *iff*

$$(\forall z_0, z_1 \in Z)(Rz_0z_1 \leftrightarrow z_0 = z_1).$$

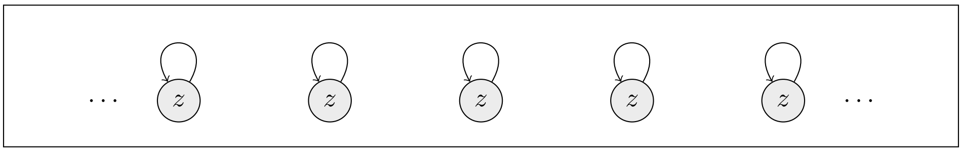

Figure 1. An idiosyncratic frame.

The strategy employed in obtaining the adequacy of the logics $S_{\blacksquare}$ – involving of course the appropriate re-definition of saturation relative to $S_{\Box}$ – yields:

[11]A set Γ of sentences is *trivial* in T iff $T \vdash \Gamma \Rightarrow \varnothing$.

Proposition 10 (Adequacy). *Let $F = (Z, R)$ be an idiosyncratic frame and S as in Definition 2. Then for any $\Gamma, \Delta \subseteq \mathcal{L}_\Box$:*

$$\Gamma \vDash^{\mathfrak{F}}_{s} \Delta \ \textit{ iff } \ S_\Box \vdash \Gamma \Rightarrow \Delta$$

holds when

- *$\mathfrak{F}$ is the class of four-valued models based on F and S is* FDE*;*
- *$\mathfrak{F}$ is the class of consistent models based on F and S is* K3*,* B3*, or* F3*;*
- *$\mathfrak{F}$ is the class of complete models based on F and S is* LP*;*
- *$\mathfrak{F}$ is the class of symmetric models based on F and S is* KS3*.*

3. Modalized truth principles

We now move on to introducing classical, non-congruent modal logics, which we will show to be the modal logics of the truth theories to be considered. These modal logics have the particular feature that their inner logic, that is, the logic governing the scope of the modal operator, will be one of the logics $S_\Box$ discussed in the previous section. After presenting these modal logics we introduce a novel version of possible world semantics in which we have both classical and nonclassical worlds. We show that our modal logics are the logics of the classical worlds of so-called mixed idiosyncratic frames.

3.1. **Axioms and rules.** We start with the most basic system, which modalizes the clauses for the positive inductive definition underlying the Kripke-Feferman approach to truth.[12] Since it will not cause any trouble in what follows, we will list the principles our classical modal logics as axioms, even though strictly speaking our logics are formulated in classical sequent calculi suitably extending the nonclassical systems of §2.

Definition 11 (Modal logic BM). *The modal logic* BM *extends classical propositional logic with:*

$(\top)$	$\Box\top$
$(\bot)$	$\neg\Box\bot$
$(\neg)$	$\Box\varphi \leftrightarrow \Box\neg\neg\varphi$
$(\wedge 1)$	$\Box(\varphi \wedge \psi) \leftrightarrow \Box\varphi \wedge \Box\psi$
$(\wedge 2)$	$\Box\neg(\varphi \wedge \psi) \leftrightarrow \Box\neg\varphi \vee \Box\neg\psi$
$(\vee 1)$	$\Box(\varphi \vee \psi) \leftrightarrow \Box\varphi \vee \Box\psi$
$(\vee 2)$	$\Box\neg(\varphi \vee \psi) \leftrightarrow \Box\neg\varphi \wedge \Box\neg\psi$
$(\Box 1)$	$\Box\varphi \leftrightarrow \Box\Box\varphi$
$(\Box 2)$	$\Box\neg\varphi \leftrightarrow \Box\neg\Box\varphi$
$(\text{FAITH}^\Box)$	$\Box\varphi \wedge \neg\Box\neg\varphi \rightarrow \varphi$

The modal logic BM$^-$ *is just like* BM *but does not assume* $(\text{FAITH}^\Box)$.

As stressed in the Introduction, the axiom $(\text{FAITH}^\Box)$ is, as we will show, the distinctive axiom of classical Kripke-style theories of truth such as the Kripke-Feferman theories. In a nutshell it asserts that if a sentence φ is nonclassically true but not false, i.e., it is not both true and false, then φ is also classically the case. Preempting our modal semantics, it asserts that if φ has a classical truth value at a

[12] As discussed in the introduction, [Fef84, §12] also considers modalizations of similar principles, but the presence of axioms of the form $\Box p \leftrightarrow p$ for atomic p of the language with $\Box$ starkly contrasts with our approach, in which we want to preserve uniform substitution at the expense of such principles for atomic formulas.

nonclassical world then it will have the same truth value at the classical worlds that see it, i.e., it expresses that the valuation is faithful with respect to classical truth values.

The logic BM can then be extended with principles corresponding to the consistency, completeness, and symmetry of $\Box$.

Definition 12 (The logics M^-, M, M^n, M^b). *The modal logic* M^n *extends* BM *with*

$$(\mathsf{D}) \qquad \neg\Box\neg\varphi \vee \neg\Box\varphi.$$

The modal logic M^b *extends* BM *with*

$$(\mathsf{D_c}) \qquad \Box\neg\varphi \vee \Box\varphi.$$

The modal logic M *(*M^-*) extend* BM *(*BM^-*) with*

$$(\mathsf{DD_c}) \qquad (\neg\Box\neg\varphi \vee \neg\Box\varphi) \vee (\Box\neg\psi \vee \Box\psi).$$

Lemma 13. *Let* M*,* M^n*,* M^b *be defined as in Definition 12. Then*

(i) $\mathsf{M} \vdash (\Box\varphi \to \varphi) \vee (\psi \to \Box\psi)$
(ii) $\mathsf{M}^\mathsf{n} \vdash \Box\varphi \to \varphi$
(iii) $\mathsf{M}^\mathsf{b} \vdash \varphi \to \Box\varphi$.

Remark 14. It's easy to see that over BM^-, $\varphi \to \Box\varphi$ entails $\mathsf{D_c}$ and $\textsc{faith}^\Box$, and $\Box\varphi \to \varphi$ entails D and $\textsc{faith}^\Box$. Proposition 24 and Corollary 29 below will entail that, in stark contrast with what happens in the truth-theoretic side, the converse implications do not hold.

Next we turn to logic whose modalities are governed by the $\mathsf{b3}$- and $\mathsf{f3}$-evaluation schemata. For notational convenience, we let

$$\begin{aligned} \Delta\varphi &:= (\Box\varphi \vee \Box\neg\varphi) \\ \nabla\varphi &:= \neg\Delta\varphi \\ \Delta(\varphi,\psi) &:= (\Delta\varphi \wedge \Delta\psi) \end{aligned}$$

Definition 15 (M^w and M^f).

(i) M^w *extends classical propositional logic with* $(\top)$, $(\bot)$, $(\neg)$, $(\wedge 1)$, $(\vee 2)$, $(\Box 1)$, $(\Box 2)$, $(\textsc{faith}^\Box)$, (D)*, and*

$$(\vee 3) \qquad \Box(\varphi \vee \psi) \leftrightarrow \Delta(\varphi,\psi) \wedge (\Box\varphi \vee \Box\psi)$$

$$(\wedge 3) \qquad \Box\neg(\varphi \wedge \psi) \leftrightarrow \Delta(\varphi,\psi) \wedge (\Box\neg\varphi \vee \Box\neg\psi)$$

We call $\mathsf{M}^{\mathsf{w}-}$ *the system* M^w *without* $(\textsc{faith}^\Box)$*.*

(ii) M^f *is formulated in* $\mathcal{L}_\Box^{\twoheadrightarrow}$ *and extends* M^w *with:*

$$(\twoheadrightarrow 1) \qquad \Box(\varphi \twoheadrightarrow \psi) \leftrightarrow \Box\neg\varphi \vee (\Box\varphi \wedge \Box\psi)$$

$$(\twoheadrightarrow 2) \qquad \Box(\neg(\varphi \twoheadrightarrow \psi)) \leftrightarrow \Box\varphi \wedge \Box\neg\psi$$

We call $\mathsf{M}^{\mathsf{f}-}$ *the system* M^f *without* $(\textsc{faith}^\Box)$*.*

Remark 16. The status of $\twoheadrightarrow$ in M^f and variants thereof is peculiar. Internally it characterises the non-material conditional of $\mathsf{F3}$, whereas externally it collapses into $\to$. It can be easily verified in fact that $(\varphi \twoheadrightarrow \psi) \leftrightarrow (\varphi \to \psi)$ is derivable in M^f-like logics, whereas from the semantics provided in the next section one can easily see that $\Box(\varphi \twoheadrightarrow \psi) \leftrightarrow \Box(\varphi \to \psi)$ is not.

The next Lemma will play a central role in what follows. It states that the derivability of sequents in the nonclassical logics of truth $S_\Box$ introduced in §2 entails the derivability of specific conditionals in the classical systems that we are currently investigating. It is in this sense that the logics $S_\Box$ are the internal logics of the modal logics introduced in this section.

LEMMA 17 (Connecting Lemma). *For* (S,T) *one of the pairs* $(\mathsf{K3},\mathsf{M^n})$, $(\mathsf{LP},\mathsf{M^b})$, $(\mathsf{KS3},\mathsf{M})$, $(\mathsf{FDE},\mathsf{BM})$, $(\mathsf{B3},\mathsf{M^w})$, $(\mathsf{F3},\mathsf{M^f})$*: if* $S_\Box \vdash \Gamma \Rightarrow \Delta$*, then* $T \vdash \bigwedge \Box\Gamma \to \bigvee \Box\Delta$.

Proof. The proof is by induction on the length of the proof in the relevant logics. Crucially, the proof for the pairs $(\mathsf{B3}_\Box,\mathsf{M^w})$ and $(\mathsf{F3},\mathsf{M^f})$ rests on the following property: for all $\varphi \in \mathcal{L}_\Box(\mathcal{L}_\Box^{\rightarrow})$, and $S \in \{\mathsf{M^w},\mathsf{M^f}\}$,

(1) $$S \vdash \nabla\varphi \text{ iff there is a } p \in \mathrm{Prop}(\varphi) \text{ such that } S \vdash \nabla p.$$

qed.

It should be noticed that, since Lemma 17 does not employ FAITH□, it can be generalized to the theories without such assumption. In the following section, it will be useful to refer directly to the *inner*, nonclassical logic of our classical modal logics of truth. The next definition makes this idea precise.

DEFINITION 18 (Inner Logic). *Given Lemma 17, we set*

$$\mathrm{I}(S) := \begin{cases} \mathsf{FDE}_\Box, & \textit{if } S = \mathsf{BM}^- \\ \mathsf{KS3}_\Box, & \textit{if } S = \mathsf{M}^- \\ \mathsf{K3}_\Box, & \textit{if } S = \mathsf{BM}^- + \mathsf{D} \\ \mathsf{LP}_\Box, & \textit{if } S = \mathsf{BM}^- + \mathsf{D_c} \\ \mathsf{B3}_\Box, & \textit{if } S = \mathsf{M^{w-}} \\ \mathsf{F3}_\Box, & \textit{if } S = \mathsf{M^{f-}} \end{cases}$$

and call $\mathrm{I}(S)$ *the inner logic* S.

3.2. **Semantics.** In this section we introduce the anticipated novel semantics for the logics described in the previous section. This amounts to considering frames endowed with classical and nonclassical worlds. In particular we are interested in what we call *mixed, idiosyncratic frames* (cf. Figure 2), that is, frames in which a classical world sees exactly one idiosyncratic nonclassical world (in the sense of Definition 9).[13]

DEFINITION 19 (Mixed idiosyncratic frame). *Let* W, Z *be disjoint nonempty sets and* $R \subseteq W \cup Z \times Z$*. A* mixed idiosyncratic frame *satisfies:*

(FUNCTIONALITY) $\forall w \in W\, \exists! v \in Z (wRv)$

(IDIOSYNCRACY) $\forall u, v \in Z (uRv \leftrightarrow u = v)$

We call a mixed idiosyncratic frame single-rooted *if* W *is a singleton.*

Mixed, idiosyncratic frames give rise to suitable models, once they are coupled with suitable valuations: in this context, a valuation takes a classical or a nonclassical world and a propositional variable and returns values in a set X with $\{0,1\} \subseteq X \subseteq \{0,1,\mathsf{b},\mathsf{n}\}$.

DEFINITION 20 (Mixed valuations). *For* (W,Z,R) *a mixed, idiosyncratic frame, a* valuation *takes a member of* $W \cup Z$ *and a* $p \in \mathrm{Prop}$ *and returns a value in* X *with* $\{0,1\} \subseteq X \subseteq \{0,1,\mathsf{b},\mathsf{n}\}$ *and* $X = \{0,1\}$ *for* $w \in W$*. In particular we call a valuation:*

- four-valued*, if* $(\forall w \in W)(\forall p \in \mathrm{Prop})(V_w(p) \in \{0,1\})$ *and* $(\forall z \in Z)(\forall p \in \mathrm{Prop})(V_z(p) \in \{0,1,\mathsf{b},\mathsf{n}\}$.
- consistent*, if* $(\forall w \in W)(\forall p \in \mathrm{Prop})(V_w(p) \in \{0,1\})$ *and* $(\forall z \in Z)(\forall p \in \mathrm{Prop})(V_z(p) \in \{1,0,\mathsf{n}\})$*;*

[13]Notice that, in Figure 2, the subframe $(\{z\},\{\langle z,z\rangle\})$ is not a mixed idiosyncratic frame in its own right.

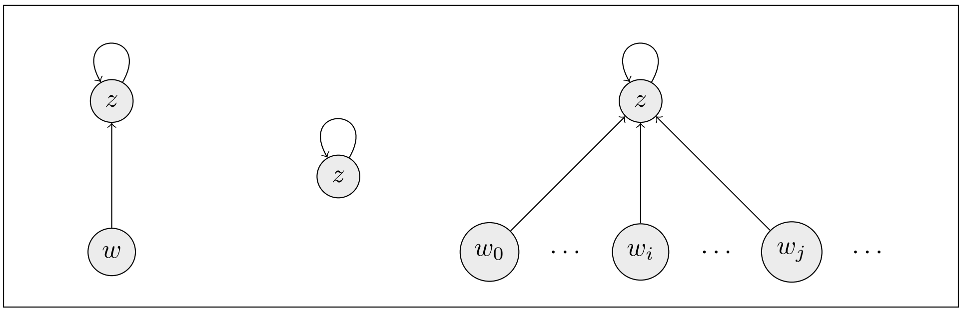

FIGURE 2. Example of a mixed idiosyncratic frame

- complete, *if* $(\forall w \in W)(\forall p \in \mathrm{Prop})(V_w(p) \in \{0,1\})$ *and* $(\forall z \in Z)(\forall p \in \mathrm{Prop})(V_z(p) \in \{1,0,\mathsf{b}\})$*;*
- symmetric*, if* $(\forall w \in W)(\forall p \in \mathrm{Prop})(V_w(p) \in \{0,1\})$ *and for all* $z \in Z$*, either* $\forall p \in \mathrm{Prop}(V_z \in \{1,0,\mathsf{b}\})$ *or* $\forall p \in \mathrm{Prop}(V_z(p) \in \{1,0,\mathsf{n}\})$ *but not both.*
- faithful*, if for all* $w \in W$*,* $z \in Z$*, and* $p \in \mathrm{Prop}$*,* Rwz *entails that* $V_w(p) = V_z(p)$ *if* $V_z(p) \in \{0,1\}$.

DEFINITION 21 (Mixed idiosyncratic model). *With* (W, Z, R) *a mixed, idiosyncratic frame and* V *a mixed valuation, the tuple* $\mathcal{M} := (W, Z, R, V)$ *is* mixed idiosyncratic model*.* $\mathcal{M}$ *is consistent (complete, symmetric, faithful), if* V *is consistent (complete, symmetric, faithful). A model is* single-rooted *if it is based on a single-rooted frame.*

The definition of truth in a mixed, idiosyncratic model combines the clauses of classical and nonclassical satisfaction.

DEFINITION 22 (Truth in mixed idiosyncratic models). *Let* (W, Z, R) *be a mixed idiosyncratic frame and* $\mathcal{M} := (W, Z, R, V)$ *a mixed idiosyncratic model and* $s \in \{\mathsf{fde}, \mathsf{k3}, \mathsf{b3}, \mathsf{f3}, \mathsf{lp}, \mathsf{ks3}\}$*. In defining truth in a model we distinguish between truth in a model at a classical world and at a nonclassical world.*

(i) Let $z \in Z$*. Then* $|\varphi|_s^{\mathcal{M},z}$ *is defined as in Definition 5. As in Definition 5 we write* $\mathcal{M}, z \Vdash_s \varphi$ *iff* $|\varphi|_s^{\mathcal{M},z} \in \{1, \mathsf{b}\}$

(ii) Let $w \in W$*. Then* $|\varphi|_s^{\mathcal{M},w}$ *is also defined using the clauses of Definition 5 unless* φ *is* $\Box\psi$*, in which case we have:*

$$|\varphi|_s^{\mathcal{M},w} = \begin{cases} 1, & \text{if } \mathcal{M}, z \Vdash_s \psi \text{, for all } z \text{ with } Rwz; \\ 0, & \text{otherwise.} \end{cases}$$

Again we write $\mathcal{M}, w \Vdash_s \varphi$ *iff* $|\varphi|_s^{\mathcal{M},w} \in \{1, \mathsf{b}\}$*. Notice that in fact a formula* φ *will not receive a nonclassical truth value at a classical world.*

Given the nature of mixed models, the definition of consequence now splits into two notions of truth preservation, one at classical worlds, and one at nonclassical worlds. The latter is simply a reformulation in the present context of the notion of consequence from Definition 6.

DEFINITION 23 (Consequence). *With* $\mathfrak{S}$ *a class of* $\mathcal{L}_\Box$*-models based on a mixed, idiosyncratic frame* (W, Z, R)*,* $s \in \{\mathsf{fde}, \mathsf{k3}, \mathsf{b3}, \mathsf{lp}, \mathsf{f3}, \mathsf{ks3}\}$*,* $\Gamma \subseteq \mathrm{Sent}_{\mathcal{L}_\Box}$*, and* $\varphi \in \mathrm{Sent}_{\mathcal{L}_\Box}$*, we let*

(i) $\Gamma \Vdash_s^{\mathfrak{S}} \varphi$ $:\Leftrightarrow$ *for all* $\mathcal{M} \in \mathfrak{S}$ *and* $w \in W$*, if* $\mathcal{M}, w \Vdash_s \gamma$ *for all* $\gamma \in \Gamma$*, then* $\mathcal{M}, w \Vdash_s \varphi$.

(ii) $\Gamma \vDash_s^{\mathfrak{S}} \Delta$ $:\Leftrightarrow$ *for all* $\mathcal{M} \in \mathfrak{S}$ *and* $z \in Z$, *if* $\mathcal{M}, z \Vdash_s \gamma$ *for all* $\gamma \in \Gamma$, *then* $\mathcal{M}, z \Vdash_s \delta$ *for all* $\delta \in \Delta$.

In the rest of the section we will prove the adequacy of our classical modal logics with respect to the semantics just introduced. As we have already seen with the notion of consequence, the adequacy theorem we are about to prove splits in two: one clause concerns the adequacy of the internal logics with respect to the nonclassical semantics of the previous section; the other clause concerns directly our classical modal logics. However, the proof of the former claim follows directly from Proposition 10. Therefore we will be mainly concerned with establishing the latter clause. The following adequacy result will not only tell us that the non-congruent modal logics we introduced are complete with respect to the classical worlds in mixed idiosyncratic frames but also that the logic that governs the transformations inside the scope of the modal operator in these logics is precisely their inner logic in the sense of Definition 18.

PROPOSITION 24 (Adequacy). *Let* $F = (W, Z, R)$ *be a mixed, idiosyncratic frame. Then for any* $\Gamma, \Delta \subseteq \mathcal{L}_\Box$ *and* $\varphi \in \mathcal{L}_\Box$, *the claims*

(i) $\Gamma \Vdash_s^{\mathfrak{F}} \varphi$ *iff* $\Gamma \vdash_S \varphi$

(ii) $\Gamma \vDash_s^{\mathfrak{F}} \Delta$ *iff* $\mathrm{I}(S) \vdash \Gamma \Rightarrow \Delta$

hold when

- $\mathfrak{F}$ *is the class of mixed, idiosyncratic models and* (s, S) *are* $(\mathsf{fde}, \mathsf{BM}^-)$;
- $\mathfrak{F}$ *is the class of mixed, idiosyncratic, symmetric models and* (s, S) *are* $(\mathsf{ks3}, \mathsf{M}^-)$;
- $\mathfrak{F}$ *is the class of mixed, idiosyncratic, consistent models and* (s, S) *are either* $(\mathsf{k3}, \mathsf{BM}^- + \mathsf{D})$, $(\mathsf{b3}, \mathsf{M}^{\mathsf{w}-})$, *or* $(\mathsf{f3}, \mathsf{M}^{\mathsf{f}-})$;
- $\mathfrak{F}$ *is the class of mixed, idiosyncratic, complete models and* (s, S) *is* $(\mathsf{lp}, \mathsf{BM}^- + \mathsf{D_c})$

A fundamental ingredient of the proof of the adequacy theorem is the definition of canonical models for our logics. Such models will reflect the mixed nature of our frames: for each logic S, classical worlds will be maximally S-consistent sets of sentences, whereas nonclassical worlds will be $\mathrm{I}(S)$-saturated sets.

DEFINITION 25 (Canonical model).

(i) *The* canonical *model for* S – $S \in \{\mathsf{BM}^-, \mathsf{M}^-, \mathsf{BM}^- + \mathsf{D}, \mathsf{BM}^- + \mathsf{D_c}, \mathsf{M}^{\mathsf{w}-}, \mathsf{M}^{\mathsf{f}-}\}$ – *is the structure* $\mathcal{M}^S := (W^S, Z^S, R^S, V^S)$, *where*

- W^S *is the set of maximally* S*-consistent sets of sentences*
- Z^S *is the set of* $\mathrm{I}(S)$*-saturated sets, where* $\mathrm{I}(S)$ *is the internal logic of* S
- $R^S(x, y)$ *is defined as follows:*

$$R^S(x,y) :\Leftrightarrow \big(x \in W^S \land \{\varphi \in \mathcal{L}_\Box \mid \Box\varphi \in x\} \subseteq y\big) \lor$$
$$\big(x \in Z^S \land \{\varphi \mid \Box\varphi \in x\} = y = \{\neg\varphi \mid \neg\Diamond\varphi \in x\}\big)$$

- $V^S := V^{\mathrm{I}(S)}$ *(cf. Definition 65)*

(ii) *As before,* (W^S, Z^S, R^S) *is the* canonical frame *for* S.

LEMMA 26 (Existence). *With* $S \in \{\mathsf{BM}^-, \mathsf{M}^-, \mathsf{BM}^- + \mathsf{D}, \mathsf{BM}^- + \mathsf{D_c}, \mathsf{M}^{\mathsf{w}-}, \mathsf{M}^{\mathsf{f}-}\}$, $v \in W^S \cup Z^S$, *and* $\varphi \in \mathcal{L}_\Box$, *if* $\Box\varphi \notin v$, *then there is a* $z \in Z^S$ *such that* $R^S vz$ *and* $\varphi \notin z$.

Proof. The proof essentially employs the Connecting Lemma (Lemma 17). The case in which $v \in Z^S$ is immediate. If $v \in W^S$, then $\neg\Box\varphi \in v$. Next, we notice that for any $\Theta \subseteq \{\psi \in \mathcal{L}_\Box \mid \Box\psi \in v\}$,

$$\mathrm{I}(S) \nvdash \Theta \Rightarrow \varphi. \tag{2}$$

In fact, if $\Theta \Rightarrow \varphi$ were derivable in $\mathrm{I}(S)$, then by Lemma 17 we would have $\bigwedge \Box\Theta \to \Box\varphi \in v$, and therefore $\Box\varphi \in v$, contradicting our assumption. By following the blueprint of Lemmata 63 and 66, we construct an $\mathrm{I}(S)$-saturated set z such that $R^S vz$ and $\varphi \notin z$. *qed.*

Lemma 27 (Truth). *With* $S \in \{\mathsf{BM}^-, \mathsf{M}^-, \mathsf{BM}^- + \mathsf{D}, \mathsf{BM}^- + \mathsf{D_c}, \mathsf{M}^{\mathsf{w}-}, \mathsf{M}^{\mathsf{f}-}\}$, $v \in W^S \cup Z^S$, *and* $\varphi \in \mathcal{L}_\Box$,

$$\varphi \in v \text{ if and only if } \mathcal{M}^S, v \Vdash_s \varphi$$

Proof. By induction on the positive complexity of φ. In the case in which φ is $\neg\Box\psi$, and $v \in W^S$, the right-to-left direction is immediate by the definition of truth and R^S. The left-to-right direction follows from Lemma 26. *qed.*

Proof of Proposition 24. The proof of (ii) essentially the same as the one of Proposition 8. For (i), the soundness direction is obtained by induction on the length of the proof in S. For the completeness direction, we first notice that, for any relevant S, the claims

(i) any S-consistent set of sentences is satisfiable in $\mathfrak{F}$

(ii) if $\Gamma \Vdash_s^{\mathfrak{F}} \varphi$ then $\Gamma \vdash_S \varphi$.

are equivalent. Now let $X \subseteq \mathcal{L}_\Box$ be S-consistent. It then suffices to find a model $\mathcal{M}$ in $\mathfrak{F}$ and a $w \in W^S$ such that $\mathcal{M}, w \Vdash \varphi$ for any $\varphi \in X$. We can then simply choose $\mathcal{M}^S$ and any $w \in W$ such that $X \subseteq w$. *qed.*

Proposition 24 yields a completeness result for the classical, non-congruent modal logics that do not assume the faithfulness axiom ($\textsc{faith}^\Box$). The axiom states that if a formula φ of $\mathcal{L}_\Box$ receives a classical truth value at a nonclassical world, then φ will have the same truth value at all classical worlds that see the nonclassical world. This informal claim is made rigorous in Lemma 28 below, which relative to mixed, idiosyncratic frames forces the valuation to be faithful in the sense of Definition 20. Lemma 28 below thus allows to transform Proposition 24 into a completeness result for the modal logics that assume ($\textsc{faith}^\Box$). The respective modal logics will be complete with respect to the class of faithful models based on mixed, idiosyncratic frames.

Lemma 28 (Faithful Models). *Let F be a mixed, idiosyncratic frame, V a mixed valuation on F and $\mathcal{M} = (F, V)$ the resulting model. Then* $\forall w \in W(\mathcal{M}, w \Vdash_s \Box\varphi \wedge \neg\Box\neg\varphi \to \varphi)$ *iff V is a faithful valuation.*

Proof. We leave it to the reader to verify that ($\textsc{faith}^\Box$) is true at all classical worlds in faithful models based on mixed, idiosyncratic frames. For the converse direction we assume for reductio that $\forall w \in W(\mathcal{M}, w \Vdash_s \Box\varphi \wedge \neg\Box\neg\varphi \to \varphi)$ on some non-faithful model based on a mixed, idiosyncratic frame, that is, for some $p \in \mathsf{Prop}$ and $w \in W$ and $z \in Z$ with Rwz: $V_z(p) \in \{0, 1\}$ but $V_w(p) \neq V_z(p)$. There are two cases:

- $V_z(p) = 1$ and $V_w(p) = 0$. Then $\mathcal{M}, w \Vdash_s \Box p \wedge \neg\Box\neg p$ and $\mathcal{M}, w \nVdash_s p$, that is, $\mathcal{M}, w \nVdash_s \Box p \wedge \neg\Box\neg p \to p$. Contradiction.
- $V_z(p) = 0$ and $V_w(p) = 1$. Then $\mathcal{M}, w \Vdash_s \Box\neg p \wedge \neg\Box\neg\neg p$ and $\mathcal{M}, w \nVdash_s \neg p$, that is, $\mathcal{M}, w \nVdash_s \Box\neg p \wedge \neg\Box\neg\neg p \to \neg p$. Contradiction.

qed.

We can now state the adequacy of the faithful modal logics, which, as we will show in §5, will serve as the modal logics of the Kripke-Feferman truth theories.

Corollary 29 (Adequacy). *Let $F = (W, Z, R)$ be a mixed, idiosyncratic frame. Then for any $\Gamma, \Delta \subseteq \mathcal{L}_\Box$ and $\varphi \in \mathcal{L}_\Box$, the claims*

(i) $\Gamma \Vdash^{\mathfrak{F}}_s \varphi$ *iff* $\Gamma \vdash_S \varphi$

(ii) $\Gamma \vDash^{\mathfrak{F}}_s \Delta$ *iff* $\mathrm{I}(S) \vdash \Gamma \Rightarrow \Delta$

holds when

- $\mathfrak{F}$ *is the class of mixed, idiosyncratic, faithful models and* (s, S) *are* $(\mathsf{fde}, \mathsf{BM})$*;*
- $\mathfrak{F}$ *is the class of mixed, idiosyncratic, symmetric models and* (s, S) *are* $(\mathsf{ks3}, \mathsf{M})$*;*
- $\mathfrak{F}$ *is the class of mixed, idiosyncratic, consistent models and* (s, S) *are either* $(\mathsf{k3}, \mathsf{M^n})$*,* $(\mathsf{b3}, \mathsf{M^w})$*, or* $(\mathsf{f3}, \mathsf{M^f})$*;*
- $\mathfrak{F}$ *is the class of mixed, idiosyncratic, complete models and* (s, S) *is* $(\mathsf{lp}, \mathsf{M^b})$*.*

When proving the truth-theoretical completeness of these modal logics, it will be useful to restrict our attention to models based on unique root models.

Corollary 30 (Single-rooted frames). *The adequacy results of Propositions 24 and 29 also hold for models based on single-rooted, mixed, idiosyncratic frames.*

Proof. The generated subframe of the canonical frame is a single-rooted mixed, idiosyncratic frame. *qed.*

4. Kripke-Feferman truth

In this section we introduce the relevant truth-theoretic background. We start with the basics of fixed-point semantics, then we introduce the base theory for our axiomatic systems of truth, and finally we define the collections of axioms of these systems.

4.1. **Peano arithmetic.** We start with an arithmetical language $\mathcal{L}_\mathbb{N}$ that includes the standard signature $\{0, \mathrm{S}, +, \times\}$, and extend it with a unary truth predicate T. We assume a canonical, monotone Gödel numbering for $\mathcal{L}_\mathrm{T}$. For e an $\mathcal{L}_\mathrm{T}$-expression, we write $\#e$ for its Gödel code and $\ulcorner e \urcorner$ for the $\mathcal{L}_\mathbb{N}$-term representing $\ulcorner e \urcorner$. $\mathcal{L}_\mathrm{T}$ features finitely many function symbols for suitable primitive recursive functions for syntactic operations on (codes of) expressions, such as:[14]

Operation	Function symbol
$\#e_1, \#e_2 \mapsto \#(e_1 = e_2)$	eq
$\#t \mapsto$ the value of the closed term $\#t$	val
$\#e \mapsto \#\neg e$	ng
$\#e_1, \#e_2 \mapsto \#(e_1 \wedge e_2)$	and
$\#e_1, \#e_2 \mapsto \#(e_1 \twoheadrightarrow e_2)$	fc
$\#v, \#\varphi \mapsto \#(\forall v \varphi)$	all
$\#t, \#\varphi(v) \mapsto \#\varphi(t/v)$	sub
$n \mapsto \#\overline{n}$	num

We assume in particular that the evaluation function is defined for the finitely many primitive recursive functions other than itself – in particular, its defining equations are part of our base theory. In this way it remains primitive recursive.

[14]In addition, we assume a function symbols for the proper subtraction function to avoid certain unintended properties of the Weak Kleene schema defined below [Spe17, CD91].

Definition 31 (Peano Arithmetic). Peano arithmetic *is the first-order system in* $\mathcal{L}_{\mathbb{N}}$ *whose axioms are:*

- $\forall x(0 \neq \mathrm{S}(x))$
- $\forall x \forall y(\mathrm{S}(x) = \mathrm{S}(y) \rightarrow x = y)$
- *the recursive equations for* $+, \times$ *and the finitely many additional primitive recursive function symbols;*
- *the axiom schema of induction:*

$$\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(\mathrm{S}x)) \rightarrow \forall x \varphi(x) \tag{$\mathsf{IND}(\mathcal{L}_{\mathbb{N}})$}$$

for all formulas $\varphi(v)$ *of* $\mathcal{L}_{\mathbb{N}}$.

Definition 32. *The system* PAT *in* $\mathcal{L}_{\mathbb{N}} \cup \{\mathrm{T}\}$ *extends the basic axioms of* PA *with all instances of induction in* $\mathcal{L}_{\mathrm{T}}$.

4.2. **Fixed-point Semantics.** Kripke-Feferman truth can be seen as axiomatizing a collection of inductive constructions of the sets of S-true sentences of $\mathcal{L}_{\mathrm{T}}$, where S is one of the nonclassical logics considered above [Mar84, Kri75, Vis89, Fef08]. Let $\mathcal{M} \vDash \mathsf{PA}$ and $\mathcal{L}_{\mathrm{T}}^{\mathcal{M}}$ be $\mathcal{L}_{\mathrm{T}}$ expanded with constants $a, b, c, \dots$ for all elements of its domain M.[15] Let True_0 be the PA-definable set of true $\mathcal{L}_{\mathbb{N}}$-equations, and False_0 be the PA-definable set of false $\mathcal{L}_{\mathbb{N}}$-equations.

We define operators on sets $S \subseteq M$ satisfying

$$\begin{aligned}(\mathcal{M}, S) \vDash \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(\mathrm{all}(v, a)) \wedge \mathrm{Cterm}_{\mathbb{N}}(b) \wedge \mathrm{Cterm}_{\mathbb{N}}(c) \wedge \mathrm{val}(b) = \mathrm{val}(c)\\ \rightarrow \big(\mathrm{Tsub}(a, v, b) \leftrightarrow \mathrm{Tsub}(a, v, c)\big)\end{aligned} \tag{REG}$$

Sets satisfying REG are called *regular* [Can89]: they simply state that the truth predicate allows for substitution of identicals.

Definition 33 (Kripke jumps). *Given some* $\mathcal{M} \vDash \mathrm{PA}$, *we are interested in two main operators on regular* $S \subseteq M$.

(i) *The* Strong-Kleene jump $\Phi \colon \mathcal{P}(M) \rightarrow \mathcal{P}(M)$ *is such that* $a \in \Phi(S)$ *if and only if*

$$\begin{aligned}&\mathcal{M} \vDash \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(a), \ \textit{and}\\ &\Big(\mathcal{M} \vDash \mathrm{True}_0(a), \ \textit{or}\\ &\mathcal{M} \vDash a = \mathrm{ng}(b) \wedge \mathrm{False}_0(b), \ \textit{or}\\ &(\mathcal{M}, S) \vDash \mathrm{Cterm}(b) \wedge a = \mathrm{sub}(\ulcorner \mathrm{T}v \urcorner, \ulcorner v \urcorner, b) \wedge \mathrm{Tval}(b), \ \textit{or}\\ &(\mathcal{M}, S) \vDash \mathrm{Cterm}(b) \wedge a = \mathrm{sub}(\ulcorner \neg \mathrm{T}v \urcorner, \ulcorner v \urcorner, b) \wedge \mathrm{Tng}(\mathrm{val}(b)), \ \textit{or}\\ &(\mathcal{M}, S) \vDash a = \mathrm{ng}(\mathrm{ng}(b)) \ \wedge \mathrm{T}(b), \ \textit{or}\\ &(\mathcal{M}, S) \vDash a = \mathrm{and}(b, c) \wedge \mathrm{T}b \wedge \mathrm{T}c, \ \textit{or}\\ &(\mathcal{M}, S) \vDash a = \mathrm{ng}(\mathrm{and}(b, c)) \wedge \mathrm{Tng}(b) \vee \mathrm{Tng}(c), \ \textit{or}\\ &(\mathcal{M}, S) \vDash a = \mathrm{all}(u, b) \wedge \forall x(\mathrm{Cterm}(x) \rightarrow \mathrm{Tsub}(b, u, x)), \ \textit{or}\\ &(\mathcal{M}, S) \vDash a = \mathrm{ng}(\mathrm{all}(u, b)) \wedge \exists x(\mathrm{Cterm}(x) \wedge \mathrm{Tsub}(\mathrm{ng}(b), u, x))\Big).\end{aligned}$$

(ii) *Let* $\mathrm{D}(x) :\leftrightarrow (\mathrm{T}x \vee \mathrm{Tng}(x))$. *The* Weak-Kleene jump $\Psi \colon \mathcal{P}(M) \rightarrow \mathcal{P}(M)$ *replaces in the definition of* Φ *the clauses for the negated conjunction and quantifiers with:*

$$\begin{aligned}&\mathcal{M} \vDash \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(a), \ \textit{and}\\ &(\mathcal{M}, S) \vDash a = \mathrm{ng}(\mathrm{and}(b, c)) \wedge (\mathrm{Tng}(b) \vee \mathrm{Tng}(c)) \wedge \mathrm{D}(a) \wedge \mathrm{D}(b) \ \textit{or}\end{aligned}$$

[15]The language expansion is not needed in the case of the standard model $\mathbb{N}$, which contains names for all natural numbers.

$$(\mathcal{M},S)\vDash a=\mathrm{all}(u,b)\wedge\forall x\,\mathrm{D}(\mathrm{sub}(b,u,x))\wedge$$
$$\forall x(\mathrm{Cterm}(x)\to\mathrm{Tsub}(b,u,x)),\ \textit{or}$$
$$(\mathcal{M},S)\vDash a=\mathrm{ng}(\mathrm{all}(u,b))\wedge\forall x\,\mathrm{D}(\mathrm{sub}(b,u,x))\wedge$$
$$\exists x(\mathrm{Cterm}(x)\wedge\mathrm{Tsub}(\mathrm{ng}(b),u,x))\Big).$$

(iii) *The* Aczel-Feferman *jump* $\Xi\colon\mathcal{P}(M)\to\mathcal{P}(M)$ *is then defined for formulas of* $\mathcal{L}_{\mathrm{T}}^{\to}$ *and follows the blueprint of the definition of* Ψ *modulo replacing* $\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}$ *with* $\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}^{\to}}$ *and the addition, to the second main conjunct, of the disjuncts:*

$$(\mathcal{M},S)\vDash a=\mathrm{fc}(b,c)\wedge\mathrm{Tng}(b)\vee(\mathrm{T}b\wedge\mathrm{T}c)$$
$$(\mathcal{M},S)\vDash a=\mathrm{ng}(\mathrm{fc}(b,c))\wedge\mathrm{T}b\wedge\mathrm{Tng}(c)$$

The subsets of $\mathrm{Sent}^{\mathcal{M}}_{\mathcal{L}_{\mathrm{T}}}:=\{a\in M\mid\mathcal{M}\vDash\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(a)\}$ and $\mathrm{Sent}^{\mathcal{M}}_{\mathcal{L}_{\mathrm{T}}^{\to}}:=\{a\in M\mid\mathcal{M}\vDash\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}^{\to}}(a)\}$ satisfying REG form a complete lattice and the operator $\Phi(\cdot)$ is monotone, therefore, by the Tarski-Knaster theorem [Tar55]:

LEMMA 34. *The operators* Φ,Ψ,Ξ *each give rise to a complete lattice of fixed points with minimal and maximal elements the sets obtained by iterating the operators on* $\varnothing$ *and on the set of sentences of the relevant language respectively.*

In what follows, when referring to a fixed point, we will always refer to a fixed point of Φ,Ψ,Ξ. Any fixed point X will have the property that: $\varphi\in X$ iff $\mathrm{T}\ulcorner\varphi\urcorner\in X$ for any sentence φ of $\mathcal{L}_{\mathrm{T}}$, where the bi-conditional is necessarily metatheoretic. This property approximates the naïve truth schema and, since $(\mathrm{T}\ulcorner\varphi\urcorner\leftrightarrow\varphi)\in X$ for $\varphi\in\mathcal{L}_{\mathbb{N}}$, it improves on the standard Tarskian solutions [Tar56] in a substantial way. This partially explains why this semantic construction is the basis of several contemporary approaches to the Liar paradox.

Different such approaches often diverge on which class of fixed points they accept.

DEFINITION 35. *A fixed point* X *on* $\mathcal{M}$ *is called:*

- consistent*, if there is no* $a\in\mathrm{Sent}^{\mathcal{M}}_{\mathcal{L}_{\mathrm{T}}}$ *(*$\mathrm{Sent}^{\mathcal{M}}_{\mathcal{L}_{\mathrm{T}}^{\to}}$*) such that* $a\in X$ *and* $\mathrm{ng}^{\mathcal{M}}(a)\in X$*;*
- complete*, if for all* $a\in\mathrm{Sent}^{\mathcal{M}}_{\mathcal{L}_{\mathrm{T}}}$ *(*$\mathrm{Sent}^{\mathcal{M}}_{\mathcal{L}_{\mathrm{T}}^{\to}}$*), either* $a\in X$ *or* $\mathrm{ng}^{\mathcal{M}}(a)\in X$.

It can be easily verified that the least fixed points are consistent, and the greatest ones are complete. It follows from the definitions that there will also be fixed points $X\in\mathfrak{L}$ that are neither consistent nor complete.

By the Diagonal Lemma, for any $\mathcal{M}\vDash\mathsf{PA}$, fixed point $X\subseteq M$, and any $i\in\omega$, we can find sentences $\tau(\bar{i},\vec{x})$ (truth-teller sentences) such that

$$\mathcal{M}\vDash\mathrm{T}\ulcorner\tau(\bar{i},x_1,\dots,x_n)\urcorner\leftrightarrow\tau(\bar{i},x_1,\dots,x_n).\tag{3}$$

In the following we use $\tau_i(\vec{x})$ as short for $\tau(\bar{i},x_1,\dots,x_n)$. The analysis of paradoxicality in [Kri75] revealed that truth-teller sentences are free to assume different truth-values in different fixed-points of Φ,Ψ,Ξ.[16]

LEMMA 36. *Let* $\mathcal{M}\vDash\mathsf{PA}$*. For* $i\in\omega$ *and any* $Y,Z\subseteq\omega$ *there is a fixed point* $X\subseteq\mathrm{Sent}^{\mathcal{M}}_{\mathcal{L}_{\mathrm{T}}}$ *(*$\mathrm{Sent}^{\mathcal{M}}_{\mathcal{L}_{\mathrm{T}}^{\to}}$*) such that*

(i) $$\#\tau_i\in X\ \textit{iff}\ i\in Y;$$

[16] The case of Weak Kleene is as usual a bit more complex: if one lacks the means for direct self-reference such as a primitive substitution function, one may not be free in assigning arbitrary truth values to truth-tellers. If for instance truth-tellers are obtained by means of existential quantification, the existence of a non determinate instance would render the quantification non determinate [CD91]. In our case we assume the means for direct self-reference and sidestep these subtle issues.

(ii) $$\#\neg\tau_i \in X \ \textit{ iff } \ i \in Z.$$

If $Y \cap Z = \emptyset$ $[Y \cup Z = \omega]$ *then* X *can be chosen to be consistent* [*complete*].

4.3. **Axioms and Rules.** In this paper Kripke-Feferman theories of truth are extensions of PA by a finite collection of axioms for the truth predicate, and, possibly, additional instances of induction in $\mathcal{L}_{\mathrm{T}}$. The truth axioms are required to be sound with respect to the fixed-point semantics just introduced and to have the additional feature that, for $\mathcal{M} \vDash \mathsf{PA}$ and A a the conjunction of such axioms:

$$X \subseteq M \text{ is a fixed point iff } (\mathcal{M}, X) \vDash \mathsf{PA} + A.$$[17]

We call KF the most basic system, whose truth predicate does not rule out interpretation in which the truth predicate is both partial and inconsistent.

Definition 37 (KF). KF *extends* PAT *with the axioms:*

(KF1) $\forall x, y\big(\mathrm{Cterm}(x) \wedge \mathrm{Cterm}(y) \to (\mathrm{Teq}(x,y) \leftrightarrow \mathrm{val}(x) = \mathrm{val}(y))\big)$

(KF2) $\forall x, y\big(\mathrm{Cterm}(x) \wedge \mathrm{Cterm}(y) \to (\mathrm{Tng}(\mathrm{eq}(x,y)) \leftrightarrow \mathrm{val}(x) \neq \mathrm{val}(y))\big)$

(KF3) $\forall x(\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(x) \to (\mathrm{Tng}(\mathrm{ng}(x)) \leftrightarrow \mathrm{T}x))$

(KF4) $\forall x \forall y(\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(\mathrm{and}(x,y)) \to (\mathrm{Tand}(x,y) \leftrightarrow \mathrm{T}x \wedge \mathrm{T}y))$

(KF5) $\forall x \forall y(\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(\mathrm{and}(x,y)) \to (\mathrm{Tng}(\mathrm{and}(x,y)) \leftrightarrow \mathrm{Tng}(x) \vee \mathrm{Tng}(y)))$

(KF6) $\forall u \forall x\big(\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(\mathrm{all}(u,x)) \to (\mathrm{Tall}(u,x) \leftrightarrow \forall y \mathrm{Tsub}(x,u,\mathrm{num}(y)))\big)$

(KF7) $\forall u \forall x\big(\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(\mathrm{all}(u,x)) \to (\mathrm{Tng}(\mathrm{all}(u,x)) \leftrightarrow \exists y \mathrm{Tsub}(\mathrm{ng}(x),u,\mathrm{num}(y)))\big)$

(KF8) $\forall x(\mathrm{Cterm}(x) \to (\mathrm{Tsub}(\ulcorner \mathrm{T}v \urcorner, \ulcorner v \urcorner, x) \leftrightarrow \mathrm{Tval}(x)))$

(KF9) $\forall x\big(\mathrm{Cterm}(x) \to$
$(\mathrm{Tsub}(\ulcorner \neg\mathrm{T}v \urcorner, \ulcorner v \urcorner, x) \leftrightarrow (\mathrm{Tng}(\mathrm{val}(x)) \vee \neg\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(\mathrm{val}(x))))\big)$

Stronger systems are obtained by adding to KF some axioms forcing a consistent or a complete truth predicate:

(CN) $$\forall x\big(\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(x) \to (\mathrm{Tng}(x) \to \neg\mathrm{T}x)\big)$$

(CM) $$\forall x\big(\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(x) \to (\neg\mathrm{T}x \to \mathrm{Tng}(x))\big)$$

The next system, called WKFC from 'Weak-Kleene Kripke Feferman with consistency', is based on a modfication of the truth clauses for connectives and quantifiers inspired to B3. We abbreviate

$$\mathrm{D}(x) :\leftrightarrow \mathrm{T}x \vee \mathrm{Tng}(x)$$
$$\mathrm{D}(x,y) :\leftrightarrow \mathrm{D}(x) \wedge \mathrm{D}(y)$$
$$\mathrm{D}^1(x) :\leftrightarrow \forall y(\mathrm{Tsub}(x, \mathrm{num}(y)) \vee \mathrm{Tsub}(\mathrm{ng}(x), u, \mathrm{num}(y)))$$

Definition 38 (WKFC). *The system* WKFC *extends* PAT *with* KF*1-4,*KF*6,*KF*8-9,* CN*, and*

(WKFC$\wedge$)
$\forall x \forall y(\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(\mathrm{and}(x,y)) \to (\mathrm{Tng}(\mathrm{and}(x,y))) \leftrightarrow \mathrm{D}(x,y) \wedge (\mathrm{Tng}(x) \vee \mathrm{Tng}(y))))$

[17] We notice that this criterion for defining Kripke-Feferman systems of truth is more selective than the $\mathbb{N}$-categoricity criterion from [FHKS15]. In fact, $\mathbb{N}$-categoricity criterion would consider a schematic version of the theories considered below – i.e. where the compositional clauses are given in schematic and not universally quantified form –, or disquotational theories in the style of PUTB (see [Hal14]) in a negation-free language, as axiomatization of suitable Kripkean fixed points.

Our truth-theoretic completeness results clearly extends to schematic, compositional versions of Kripke-Feferman systems, and it should easily extend to suitable disquotational systems.

(WKFC∀)
$\forall u \forall x \big(\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}(\mathrm{all}(u,x)) \to (\mathrm{Tng}(\mathrm{all}(u,x)) \leftrightarrow \mathrm{D}^1(x) \wedge \exists y \mathrm{Tsub}(\mathrm{ng}(x), u, \mathrm{num}(y)))\big)$

The last collection of axioms results in the system DT introduced by Feferman in [Fef08] – and also Feferman's preferred system of truth – whose truth predicate is based in the logic F3.[18]

Definition 39 (DT). *The system* DT *extends* WKFC *with*

(DT↠) $\qquad \forall x \forall y (\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}^{\twoheadrightarrow}}(\mathrm{fc}(x,y)) \to (\mathrm{Tfc}(x,y) \leftrightarrow (\mathrm{Tng}(x) \vee (\mathrm{T}x \wedge \mathrm{T}y))))$

(DT¬↠) $\qquad \forall x \forall y (\mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}^{\twoheadrightarrow}}(\mathrm{fc}(x,y)) \to (\mathrm{Tng}(\mathrm{fc}(x,y)) \leftrightarrow (\mathrm{T}x \wedge \mathrm{Tng}(y))))$

Remark 40. Similarly to what is described by Remark 16, $\to$ and $\twoheadrightarrow$ are externally equivalent in DT and variants thereof, whereas by the properties of the fixed-point semantics introduced $\mathrm{T}\ulcorner \varphi \to \psi \urcorner \leftrightarrow \mathrm{T}\ulcorner \varphi \twoheadrightarrow \psi \urcorner$ is *not* provable in such theories.

The next lemma collects some simple facts concerning the provability and unprovability of Liar sentences in the Kripke-Feferman systems just introduced.

Lemma 41. *Let l be a $\mathcal{L}_{\mathbb{N}}$ term such that* PAT $\vDash l = \ulcorner \neg \mathrm{T} l \urcorner$*, and let* $\lambda :\leftrightarrow \neg \mathrm{T} l$*. We have:*

(i) $\Sigma \nvdash \lambda$, $\Sigma \nvdash \neg\lambda$ *for* $\Sigma \in \{\mathsf{KF}, \mathsf{KF} + \mathsf{CN} \vee \mathsf{CM}\}$*;*
(ii) $\Sigma \vdash \lambda \wedge \neg \mathrm{T}\ulcorner \neg\lambda \urcorner \wedge \neg \mathrm{T}\ulcorner \lambda \urcorner$*, for* $\Sigma \in \{\mathsf{KF} + \mathsf{CN}, \mathsf{WKFC}, \mathsf{DT}\}$*;*
(iii) $\mathsf{KF} + \mathsf{CM} \vdash \mathrm{T}\ulcorner \lambda \urcorner \wedge \neg\lambda$*.*

5. Truth-theoretic Completeness

In this section we establish the fundamental link between the classical modal logics introduced in §3.1 and the Kripke-Feferman theories of truth in the form of Solovay-style completeness results.

5.1. The modal logic of KF.

We start by establishing the truth-theoretic completeness of the basic Kripke-Feferman system KF.

Definition 42 (Truth-realization, Truth-interpretation). *A* truth-realization *is a function* $\star\colon \mathrm{Prop} \to \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}$*. Each realization induces a* truth-interpretation*, i.e. a function* $\mathfrak{I}^\star \colon \mathrm{Sent}_{\mathcal{L}_{\Box}} \to \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}$ *such that:*

$$\mathfrak{I}^\star(\varphi) = \begin{cases} p_i^\star, & \text{if } \varphi := p_i \\ 0 = 0, & \text{if } \varphi = \top \\ 0 = 1, & \text{if } \varphi = \bot \\ \neg \mathfrak{I}^\star(\psi), & \text{if } \varphi = \neg\psi \\ \mathfrak{I}^\star(\psi) \wedge \mathfrak{I}^\star(\chi), & \text{if } \varphi = \psi \wedge \chi \\ \mathrm{T}\,(\ulcorner \mathfrak{I}^\star(\psi) \urcorner), & \text{if } \varphi = \Box\psi. \end{cases}$$

The definition can easily be extended to the case of $\mathcal{L}_{\Box}^{\twoheadrightarrow}$ and $\mathcal{L}_{\mathrm{T}}^{\twoheadrightarrow}$ by adding an extra clause for the truth interpretation:

$$\mathfrak{I}^\star(\psi) \twoheadrightarrow \mathfrak{I}^\star(\chi), \quad \text{if } \varphi = \psi \twoheadrightarrow \chi$$

Since this will be clear from the context, we will use the term truth-interpretation *for both translations.*

The following is the main result of the present subsection, and establishes that BM is the modal logic of the basic system KF.

[18] To be precise, we are not presenting here the original axiomatization by Feferman, but a variant of it considered in [Fuj10].

THEOREM 43. *For all* $\varphi \in \mathrm{Sent}_{\mathcal{L}_\Box}$, $\mathsf{BM} \vdash \varphi$ *if and only if for all realizations* $\star$, $\mathsf{KF} \vdash \mathfrak{I}^\star(\varphi)$.

The proof of theorem 43 consists of two parts: the soundness and the completeness of BM. The soundness direction is established via a straightforward induction on the length of the proof in BM.

LEMMA 44 (Soundness of BM). *For all* $\varphi \in \mathrm{Sent}_{\mathcal{L}_\Box}$, *if* $\mathsf{BM} \vdash \varphi$, *then for all realizations* $\star$, $\mathsf{KF} \vdash \mathfrak{I}^\star(\varphi)$.

The converse direction will be proven in its contrapositive form.

LEMMA 45 (Truth-Theoretic Completeness). *For all* $\varphi \in \mathrm{Sent}_{\mathcal{L}_\Box}$, *if* $\mathsf{BM} \nvdash \varphi$, *then there exists a realization* $\star$, $\mathsf{KF} \nvdash \mathfrak{I}^\star(\varphi)$.

Before proving Lemma 45, we briefly sketch the general proof strategy. We start the proof by assuming $\mathsf{BM} \nvdash \varphi$. By the modal completeness theorem and, in particular, Corollaries 29 and 30, we know that there is a mixed, faithful single rooted idiosyncratic FDE-model $\mathcal{M}$ such that at its classical root w, $\mathcal{M}, w \Vdash_{\mathsf{fde}} \neg\varphi$. We then choose a particular truth-realization, which allows us to "code up" or "mimic" the valuation of the FDE-model by choosing an appropriate KF-model. In the KF-model the truth interpretation of $\neg\varphi$ will be true. Hence the truth-interpretation of φ under the chosen truth-realization, which we label the WITNESS REALIZATION, is not a theorem of KF.

The idea behind the WITNESS REALIZATION is to employ the special properties of truth-tellers to interpret propositional variables. In particular, as we have seen in Lemma 36, for any collection of truth-tellers we can find KF-models that declare truth tellers of the particular collection true (false). Hence, truth-tellers display the right amount of freedom required for interpreting propositional variables of the modal language: truth-tellers behave like propositional variables over the lattice of fixed-point models.[19] However, the WITNESS REALIZATION encodes the interpretation of propositional variables across different worlds, a classical and a nonclassical world, which requires interpreting a propositional variables as a conjunction of a truth-teller and a negated truth-teller. The following Lemma, which follows easily from Lemma 36, can be seen as a particular application of this fact.

LEMMA 46. *Let* Φ *be as above. Then for any mixed, faithful, single-rooted idiosyncratic model* $\mathcal{M}$ *based on an evaluation scheme* $e \in \{\mathsf{fde}, \mathsf{ks3}, \mathsf{k3}, \mathsf{lp}\}$ *and any* $\mathcal{N} \models \mathsf{PA}$ *we can find a fixed point* $S \subset N$ *(*N *being the domain of* $\mathcal{N}$*) of* Φ *such that for all* $p_j \in \mathrm{Prop}$ *with* $j \in \omega$*:*

(i) $\#\tau_{2j} \in S$ *iff* $\mathcal{M}, w \Vdash_e p_j$ *or* $\mathcal{M}, z \Vdash_e p_j$*;*
(ii) $\#\neg\tau_{2j} \in S$ *iff* $\mathcal{M}, z \Vdash_e \neg p_j$*;*
(iii) $\#\tau_{2j+1} \in S$ *iff* $\mathcal{M}, w \Vdash_e \neg p_j$ *and* $\mathcal{M}, z \Vdash_e p_j$*;*
(iv) $\#\neg\tau_{2j+1} \in S$ *iff* $\mathcal{M}, z \Vdash_e p_j$*.*

Moreover, for k3 *(*lp*) we can find a consistent (complete) fixed point* S*; for* ks3 *depending on the model we can choose either a consistent or a complete fixed point* S*.*

DEFINITION 47 (Witness Realization). *Let* $\bullet\colon \mathrm{Prop} \to \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}$ *be a truth realization such that for all* $j \in \omega$

$$p_j^\bullet = \tau_{2j} \wedge \neg\tau_{2j+1}.$$

$\bullet$ *is called the* WITNESS REALIZATION.

[19]Moreover, truth-tellers are essentially obtained by diagonalization. This reveals an interesting analogy with Solovay's original argument for the completeness of provablity logic, in which diagonalization is also essentially employed – however, Fedor Pakhomov has recently provided a diagonalization-free proof [Pak17]. We thank an anonymous referee for pointing out this analogy.

Our next claim is key for the proof of the main lemma of this section, and describes the behaviour of the witness realization at the nonclassical world.

LEMMA 48. *$\mathcal{M}$, S and e be as in Lemma 46 and $\bullet$ be the* WITNESS REALIZATION. *Then for all $\varphi \in \mathcal{L}_\Box$:*

(i) *if $\mathcal{M}, z \Vdash_e \varphi$, then $\mathfrak{I}^\bullet(\varphi) \in S$;*
(ii) *if $\mathcal{M}, z \nVdash_e \varphi$, then $\mathfrak{I}^\bullet(\varphi) \notin S$.*

Proof. Both cases are proved by an induction on the positive complexity of φ. We discuss the base cases, the remaining cases are easily obtained by induction hypothesis. We start with item (i). Suppose $\varphi = p_j$ for some $j \in \omega$ and $\mathcal{M}, z \Vdash_e \varphi$. Then, by Lemma 46(i), $\#\tau_{2j} \in S$ and, by (iv), $\#\neg\tau_{2j+1} \in S$. Since S is a fixed point of Φ, this implies that $\#(\tau_{2j} \wedge \neg\tau_{2j+1}) \in S$, i.e., $\#p_j^\bullet \in S$. Similarly, if $\varphi = \neg p_j$ for some $j \in \omega$ and $\mathcal{M}, z \Vdash_e \varphi$, then by Lemma 46(ii) $\#\neg\tau_{2j} \in S$. From this we may conclude that $\#(\neg\tau_{2j} \vee \neg\neg\tau_{2j+1}) \in S$, that is $\#\neg(\tau_{2j} \wedge \neg\tau_{2j+1}) \in S$. The latter is just $\#\mathfrak{I}^\bullet(\neg p_j) \in S$.

For item (ii) we assume $\varphi = p_j$ for some $j \in \omega$ and $\mathcal{M}, z \nVdash_e \varphi$. Then by Lemma 46 (iv) $\#\neg\tau_{2j+1} \notin S$ which implies that $\#(\tau_{2j} \wedge \neg\tau_{2j+1}) \notin S$, i.e., $\#\mathfrak{I}^\bullet(p_j) \notin S$. We now assume $\varphi = \neg p_j$ for some $j \in \omega$ and $\mathcal{M}, z \nVdash_e \varphi$. By Lemma 46 (ii) we infer $\#\neg\tau_{2j} \notin S$. Now, we distinguish between two cases: in the first case $\mathcal{M}, z \nVdash_e p_j$ and by (iii) $\#\tau_{2j+1} \notin S$. Alternatively, $\mathcal{M}, z \Vdash_e p_j$ but then, since we are working in a faithful model $\mathcal{M}, w \Vdash_e p_j$, i.e., $\mathcal{M}, w \nVdash_e \neg p_j$ and again by (iii) it follows that $\#\tau_{2j+1} \notin S$. We can conclude that $\#\neg(\tau_{2j} \wedge \neg\tau_{2j+1}) \notin S$. But the latter is just $\#\mathfrak{I}^\bullet(\neg p_j) \notin S$.

qed.

We can then establish the main lemma to the truth-completeness of KF.

LEMMA 49 (Main Lemma). *Let $\mathcal{M}$, S and e be as in Lemma 46 and $\bullet$ be the* WITNESS REALIZATION. *Then, for all $\varphi \in \mathcal{L}_\Box$: if $\mathcal{M}, w \Vdash_e \varphi$, then $(\mathcal{N}, S) \vDash \mathfrak{I}^\bullet(\varphi)$.*

Proof. The proof is again by induction on the positive complexity of φ. We cover the base cases and the cases of the modal operator. The induction step for the remaining operators and quantifiers is immediate by the properties of KF-models. We assume $\varphi = p_j$ and $\mathcal{M}, w \Vdash_e \varphi$. We know by Lemma 46 (i) that $\#\tau_{2j} \in S$ and by (iii) that $\#\tau_{2j+1} \notin S$. By the properties of truth tellers this implies $(\mathcal{N}, S) \vDash \mathfrak{I}^\bullet(p_j)$.

We now assume $\varphi = \neg p_j$ and $\mathcal{M}, w \Vdash_e \varphi$. We distinguish between case (a) $\mathcal{M}, z \nVdash_e p_j$ and case (b) where $\mathcal{M}, z \Vdash_e p_j$. In case (a) we can infer Lemma 46 (i) that $\#\tau_{2j} \notin S$, which suffices to show that $(\mathcal{N}, S) \vDash \mathfrak{I}^\bullet(\neg p_j)$. In case (b) we infer by 46 (iii) that $\#\tau_{2j+1} \in S$, which again suffices to show that $(\mathcal{N}, S) \vDash \mathfrak{I}^\bullet(\neg p_j)$.

Let $\varphi = \Box\psi$ and $\mathcal{M}, w \Vdash_e \varphi$. We know that $\mathcal{M}, w \Vdash_e \varphi$ if and only if $\mathcal{M}, z \Vdash_e \psi$. But from $\mathcal{M}, z \Vdash_e \psi$ we infer by Lemma 48 that $\#\mathfrak{I}^\bullet(\psi) \in S$. Thus $(\mathcal{N}, S) \vDash \mathrm{T}\ulcorner\mathfrak{I}^\bullet(\psi)\urcorner$, which by Definition 42 is just $(\mathcal{N}, S) \vDash \mathfrak{I}^\bullet(\varphi)$.

Finally, let $\varphi = \neg\Box\psi$ and $\mathcal{M}, w \Vdash_e \varphi$. We know that $\mathcal{M}, w \Vdash_e \varphi$ if and only if $\mathcal{M}, z \nVdash_e \psi$. But from $\mathcal{M}, z \nVdash_e \psi$ we infer by Lemma 48 that $\#\mathfrak{I}^\bullet(\psi) \notin S$. Thus $(\mathcal{N}, S) \vDash \neg\mathrm{T}\ulcorner\mathfrak{I}^\bullet(\psi)\urcorner$, which by Definition 42 is just $(\mathcal{N}, S) \vDash \mathfrak{I}^\bullet(\varphi)$.

qed.

We can now prove the truth-completeness of BM.

Proof of Lemma 45. Assume BM $\nvdash \varphi$. Then by Corollaries 29 and 30 we know that there is a mixed, faithful single-rooted idiosyncratic model $\mathcal{M}$ such that at its root w, $\mathcal{M}, w \Vdash_{\mathsf{fde}} \neg\varphi$. By Lemma 46 we then choose an appropriate fixed-point model $(\mathcal{N}, S)$ of KF – e.g. a fixed-point model based on $\mathbb{N}$ – such that by the Main Lemma, i.e. Lemma 49, $(\mathcal{N}, S) \models \mathfrak{I}^\bullet(\neg\varphi)$, where $\bullet$ is the WITNESS REALIZATION.

The latter implies $\mathsf{KF} \nvdash \mathfrak{I}^{\bullet}(\varphi)$ and hence that there is truth-realization $\star$ such that $\mathsf{KF} \nvdash \mathfrak{I}^{\star}(\varphi)$. *qed.*

Proof of Theorem 43. By Lemma 44 and Lemma 45. *qed.*

Corollary 50. *For all $\varphi \in \mathrm{Sent}_{\mathcal{L}_\Box}$*

(ks3) $\mathsf{M} \vdash \varphi$ *if and only if for all realizations* $\star\,(\mathsf{KF} + \mathsf{CM} \vee \mathsf{CN} \vdash \mathfrak{I}^{\star}(\varphi))$;

(k3) $\mathsf{M^n} \vdash \varphi$ *if and only if for all realizations* $\star\,(\mathsf{KF} + \mathsf{CN} \vdash \mathfrak{I}^{\star}(\varphi))$;

(lp) $\mathsf{M^b} \vdash \varphi$ *if and only if for all realizations* $\star\,(\mathsf{KF} + \mathsf{CM} \vdash \mathfrak{I}^{\star}(\varphi))$.

Proof. The soundness of M ($\mathsf{M^n}$, $\mathsf{M^b}$) with respect to $\mathsf{KF} + \mathsf{CM} \vee \mathsf{CN}$ ($\mathsf{KF} + \mathsf{CN}$, $\mathsf{KF} + \mathsf{CM}$) follows again by an induction on the length of a proof in M ($\mathsf{M^n}, \mathsf{M^b}$). For the converse direction, i.e. the truth-theoretic completeness, we adopt the strategy employed in proving Lemma 45: we assume that some formula φ is not provable in the modal logic at stake. We then apply the modal completeness theorem to find a suitable faithful, mixed single rooted idiosyncratic model that falsifies φ. Then using Lemma 46 we can find suitable fixed-point models of $\mathsf{KF} + \mathsf{CM} \vee \mathsf{CN}$ ($\mathsf{KF} + \mathsf{CN}$, $\mathsf{KF} + \mathsf{CM}$) in which the truth-interpretation based on our Witness Realization of φ is false. *qed.*

Remark 51 (Uniform Completeness). We note that in the proof of Lemma 45 the Witness Realization is kept fixed for all $\varphi \in \mathsf{Sent}_{\mathcal{L}_\Box}$. This entails that our result directly yields a *uniform completeness* result as a corollary: there exists a single realization $\bullet$ such that for all $\varphi \in \mathsf{Sent}_{\mathcal{L}_\Box}$, if the modal logic does not prove φ, then $\mathfrak{I}^{\bullet}(\varphi)$ is not provable in the corresponding truth theory.

Finally, before moving to the next section, in which we consider strengthenings of some of the claims just obtained, we notice that Lemma 48 provides also a direct proof of the truth-theoretic completeness of the modal logics $S_\Box$ with the respects to the corresponding – in the sense of the underlying nonclassical logics – nonclassical axiomatizations of Kripkean truth in the style of PKF from [HH06].[20]

5.2. **The modal logics of WKFC, DT, and of Kripke's fixed points.** In this section we determine the modal logics of the truth-theories based on Weak Kleene. In doing so, we will use an alternative argument to the one employed in the previous subsection, that will also deliver alternative truth-theoretic completeness proofs for $\mathsf{KF}+\mathsf{CN}$ and $\mathsf{KF}+\mathsf{CM}$. However, such alternative strategies are not different proofs of the same results, but in fact yield much stronger claims, namely they determine the modal logic of *all consistent extensions* (not necessarily recursively enumerable) of the truth systems considered. For example, [Bur14] proposed an extension of KF in $\mathcal{L}_{\mathrm{T}}$ with a minimality schema – called $\mathsf{KF}\mu$ –, that was intended to axiomatize Kripke's minimal fixed point model. Our result will show that $\mathsf{M^n}$ is the modal logic not only of $\mathsf{KF}\mu$, but also to stronger extensions of KF such as the set of sentences of $\mathcal{L}_{\mathrm{T}}$ satisfied in the model $(\mathbb{N}, \mathcal{I}_\Phi)$, where $\mathcal{I}_\Phi$ is the minimal fixed point of Φ.

Theorem 43 establishes that BM is the modal logic of KF. However, KF is not the only first-order theory whose modal logic is BM.

Observation 1. There are $2^{\aleph_0}$ recursive, consistent, and mutually inconsistent extensions of KF in $\mathcal{L}_{\mathrm{T}}$ whose modal logic is BM.

Proof. By Gödel's incompleteness theorem, one constructs a copy of the full binary tree starting with KF: each node, say labelled with T, has two children labelled with $T + \gamma_T$ and $T + \neg\gamma_T$, where γ_T is an arithmetical sentence undecidable in T.

[20]For formulations of PKF in the various nonclassical logics discussed in the paper see [Nic18, CS20].

Each node will then be consistent (assuming KF is), and inconsistent with the other non-root nodes.

The strategy leading to Theorem 43 can then be employed to show that BM is the modal logic of the resulting theories. It is worth noting that for arithmetically unsound theories the required countermodel in the proof of theorem 43 can only be nonstandard. *qed.*

At the same time, Corollary 50 shows the existence of consistent, recursively enumerable extensions of KF in $\mathcal{L}_{\mathrm{T}}$ whose modal logic is not BM. A natural question is then whether there are extensions of KF in $\mathcal{L}_{\mathrm{T}}$ whose modal logic is stable under further extensions. As anticipated, the answer turns out to be positive: to establish this, we present an alternative proof of Corollary 50. As a consequence, we will obtain a truth completeness proof for WKFC and DT which will also be stable under consistent extensions, *including the $\mathcal{L}_{\mathrm{T}}$-sentences true in fixed-point models.*

LEMMA 52.

(i) *Let $\mathcal{M} = (\{w\}, \{z\}, R, V)$ be a mixed, idiosyncratic, faithful, single-rooted, consistent model of $\mathcal{L}_{\Box}$. Then:*
 a. *For all $\varphi \in \mathcal{L}_{\Box}$, if $\mathcal{M}, w \Vdash_{\mathsf{k3}} \varphi$, then there is some truth-realization $\star$ such that $\mathsf{KF} + \mathsf{CN} \vdash \mathfrak{I}^{\star}(\varphi)$.*
 b. *For all $\varphi \in \mathcal{L}_{\Box}$, if $\mathcal{M}, w \Vdash_{\mathsf{b3}} \varphi$, then there is some truth-realization $\star$ such that $\mathsf{WKFC} \vdash \mathfrak{I}^{\star}(\varphi)$.*
 c. *For all $\varphi \in \mathcal{L}_{\Box}^{\rightarrow}$, if $\mathcal{M}, w \Vdash_{\mathsf{f3}} \varphi$, then there is some truth-realization $\star$ such that $\mathsf{DT} \vdash \mathfrak{I}^{\star}(\varphi)$.*

(ii) *Let $\mathcal{M} = (\{w\}, \{z\}, R, V)$ be a mixed, idiosyncratic, faithful, single-rooted, complete model of $\mathcal{L}_{\Box}$. Then for all $\varphi \in \mathcal{L}_{\Box}$, if $\mathsf{M}^{\mathsf{b}}, w \Vdash_{\mathsf{lp}} \varphi$, then there is some truth-realization $\star$ such that $\mathsf{KF} + \mathsf{CM} \vdash \mathfrak{I}^{\star}(\varphi)$.*

Lemma 52 relies in turn on the following, crucial Lemmata, which are stronger versions of Lemma 48.

LEMMA 53. *Let $\mathcal{M} = (\{w\}, \{z\}, R, V)$ be as in Lemma 52(i). Then there is a realization $\circ$ such that:*

(i) *for all $\varphi \in \mathcal{L}_{\Box}$,*
 a. *if $\mathcal{M}, z \Vdash_{\mathsf{k3}} \varphi$, then $\mathsf{KF} + \mathsf{CN} \vdash \mathrm{T}\ulcorner \mathfrak{I}^{\circ}(\varphi)\urcorner$;*
 b. *if $\mathcal{M}, z \nVdash_{\mathsf{k3}} \varphi$, then $\mathsf{KF} + \mathsf{CN} \vdash \neg\mathrm{T}\ulcorner \mathfrak{I}^{\circ}(\varphi)\urcorner$.*

(ii) *for all $\varphi \in \mathcal{L}_{\Box}$,*
 a. *if $\mathcal{M}, z \Vdash_{\mathsf{b3}} \varphi$, then $\mathsf{WKFC} \vdash \mathrm{T}\ulcorner \mathfrak{I}^{\circ}(\varphi)\urcorner$;*
 b. *if $\mathcal{M}, z \nVdash_{\mathsf{b3}} \varphi$, then $\mathsf{WKFC} \vdash \neg\mathrm{T}\ulcorner \mathfrak{I}^{\circ}(\varphi)\urcorner$.*

(iii) *for all for all $\varphi \in \mathcal{L}_{\Box}^{\rightarrow}$*
 a. *if $\mathcal{M}, z \Vdash_{\mathsf{f3}} \varphi$, then $\mathsf{DT} \vdash \mathrm{T}\ulcorner \mathfrak{I}^{\circ}(\varphi)\urcorner$;*
 b. *if $\mathcal{M}, z \nVdash_{\mathsf{f3}} \varphi$, then $\mathsf{DT} \vdash \neg\mathrm{T}\ulcorner \mathfrak{I}^{\circ}(\varphi)\urcorner$.*

Proof. Let $\circ$ be the realization

$$p^{\circ} = \begin{cases} 0 = 0, & \text{if } V_z(p) = 1, \\ \lambda, & \text{if } V_w(p) = 1 \text{ and } V_z(p) = \mathsf{n}, \\ \neg\lambda, & \text{if } V_w(p) = 0 \text{ and } V_z(p) = \mathsf{n}, \\ 0 = 1, & \text{otherwise.} \end{cases}$$

We verify in some detail (ii) and (iii), because (i) easily follows from the axioms of $\mathsf{KF} + \mathsf{CN}$ and Lemma 41. Both cases are inductions on the positive complexity of φ.

(ii) The noteworthy cases for a. are the cases in which φ is p or $\neg p$, in which one employs the properties of liar sentences in WKFC, and the one in which $\varphi := \neg(\psi \wedge \chi)$. Since $\mathcal{M}, z \Vdash_{\mathsf{b3}} \varphi$, there are three cases:

- $\mathcal{M}, z \Vdash_{\mathsf{b3}} \neg\psi$ and $\mathcal{M}, z \Vdash_{\mathsf{b3}} \neg\chi$
- $\mathcal{M}, z \Vdash_{\mathsf{b3}} \psi$ and $\mathcal{M}, z \Vdash_{\mathsf{b3}} \neg\chi$
- $\mathcal{M}, z \Vdash_{\mathsf{b3}} \neg\psi$ and $\mathcal{M}, z \Vdash_{\mathsf{b3}} \chi$

By induction hypothesis, in all cases on obtains in WKFC that $\mathrm{D}(\ulcorner\mathfrak{I}^{\circ}(\psi)\urcorner, \ulcorner\mathfrak{I}^{\circ}(\chi)\urcorner)$ and $\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\neg\psi)\urcorner \vee \mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\neg\chi)\urcorner$. Therefore, $\mathsf{WKFC} \vdash \mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\neg(\psi \wedge \chi))\urcorner$.

Symmetrically, b.'s atomic cases follow from the properties of the liar sentence in WKFC. For the crucial case of $\varphi := \neg(\psi \wedge \chi)$, the assumption yields two main cases: $|\varphi|^{\mathcal{M},z}_{\mathsf{b3}} = \mathsf{n}$ or $|\varphi|^{\mathcal{M},z}_{\mathsf{b3}} = \mathsf{0}$. The latter case is readily obtained: by induction hypothesis, WKFC proves $\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\psi \wedge \chi)\urcorner)$, and so $\neg\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\neg(\psi \wedge \chi))\urcorner$. For the first case,

$$|\psi|^{\mathcal{M},z}_{\mathsf{b3}} = |\neg\psi|^{\mathcal{M},z}_{\mathsf{b3}} = \mathsf{n} \text{ or } |\chi|^{\mathcal{M},z}_{\mathsf{b3}} = |\neg\chi|^{\mathcal{M},z}_{\mathsf{b3}} = \mathsf{n}.$$

If $|\psi|^{\mathcal{M},z}_{\mathsf{b3}} = \mathsf{n}$, the induction hypothesis entails that $\mathsf{WKFC} \vdash \neg\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\psi)\urcorner$ and $\mathsf{WKFC} \vdash \neg\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\neg\psi)\urcorner$. Therefore, $\mathsf{WKFC} \vdash \neg\mathrm{D}(\ulcorner\mathfrak{I}^{\circ}(\psi)\urcorner, \ulcorner\mathfrak{I}^{\circ}(\chi)\urcorner)$ and $\mathsf{WKFC} \vdash \neg\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\neg(\psi \wedge \chi))\urcorner$.

(iii) All cases are analogous to the proof of (ii) except of course the cases in which $\varphi = \psi \twoheadrightarrow \psi$ or $\varphi = \neg(\psi \twoheadrightarrow \chi)$, which we now consider.

a: if $|\psi \twoheadrightarrow \chi|^{\mathcal{M},z}_{\mathsf{f3}} = \mathsf{1}$, then either $|\psi|^{\mathcal{M},z}_{\mathsf{f3}} = \mathsf{0}$ or $|\psi|^{\mathcal{M},z}_{\mathsf{f3}} = |\chi|^{\mathcal{M},z}_{\mathsf{f3}} = 1$. In either case, by induction hypothesis and ($\mathsf{DT}\twoheadrightarrow$), we have $\mathsf{DT} \vdash \mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\psi) \twoheadrightarrow \mathfrak{I}^{\circ}(\chi)\urcorner$, and therefore $\mathsf{DT} \vdash \mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\psi \twoheadrightarrow \chi)\urcorner$. If $|\psi \twoheadrightarrow \chi|^{\mathcal{M},z}_{\mathsf{f3}} = \mathsf{0}$, then $|\psi|^{\mathcal{M},z}_{\mathsf{f3}} = 1$ and $|\chi|^{\mathcal{M},z}_{\mathsf{f3}} = 0$. The induction hypothesis yields that $\mathsf{DT} \vdash \mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\psi)\urcorner \wedge \mathrm{T}\ulcorner\neg\mathfrak{I}^{\circ}(\chi)\urcorner$, so the claim follows by ($\mathsf{DT}\neg\twoheadrightarrow$).

b: if $\mathcal{M}, z \nVdash_{\mathsf{f3}} \psi \twoheadrightarrow \chi$, then $\mathcal{M}, z \nVdash_{\mathsf{f3}} \neg\psi$ and $\mathcal{M}, z \nVdash_{\mathsf{f3}} \psi \wedge \chi$. Therefore, either $\mathcal{M}, z \nVdash_{\mathsf{f3}} \neg\psi$ and $\mathcal{M}, z \nVdash_{\mathsf{f3}} \psi$, or $\mathcal{M}, z \nVdash_{\mathsf{f3}} \neg\psi$ and $\mathcal{M}, z \nVdash_{\mathsf{f3}} \chi$. Thus, by induction hypothesis,

$$\mathsf{DT} \vdash \neg\mathrm{T}\ulcorner\neg\mathfrak{I}^{\circ}(\psi)\urcorner \wedge \neg\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\psi)\urcorner \text{ or } \mathsf{DT} \vdash \neg\mathrm{T}\ulcorner\neg\mathfrak{I}^{\circ}(\psi)\urcorner \wedge \neg\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\chi)\urcorner$$

By ($\mathsf{DT}\twoheadrightarrow$), we can in either case conclude that $\mathsf{DT} \vdash \neg\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\psi) \twoheadrightarrow \mathfrak{I}^{\circ}(\chi)\urcorner$, that is $\mathsf{DT} \vdash \neg\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\psi \twoheadrightarrow \chi)\urcorner$. Finally, if $\mathcal{M}, z \nVdash_{\mathsf{f3}} \neg(\psi \twoheadrightarrow \chi)$, then either $\mathcal{M}, z \nVdash_{\mathsf{f3}} \psi$ or $\mathcal{M}, z \nVdash_{\mathsf{f3}} \neg\chi$. Therefore, by induction hypothesis,

$$\mathsf{DT} \vdash \neg\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\psi)\urcorner \text{ or } \mathsf{DT} \vdash \neg\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\chi)\urcorner$$

In either case, by ($\mathsf{DT}\neg\twoheadrightarrow$) and the definition of $\mathfrak{I}^{\circ}$, one obtains that $\neg\mathrm{T}\ulcorner\mathfrak{I}^{\circ}(\psi \twoheadrightarrow \chi)\urcorner$ is provable in DT. *qed.*

LEMMA 54. *Let $\mathcal{M} = (\{w\}, \{z\}, R, V)$ be as in Lemma 52(ii). Then there is a realization † such that:*

(i) *if $\mathcal{M}, z \Vdash_{\mathsf{lp}} \varphi$, then $\mathsf{KF} + \mathsf{CM} \vdash \mathrm{T}\ulcorner\mathfrak{I}^{\dagger}(\varphi)\urcorner$;*

(ii) *if $\mathcal{M}, z \nVdash_{\mathsf{lp}} \varphi$, then $\mathsf{KF} + \mathsf{CM} \vdash \neg\mathrm{T}\ulcorner\mathfrak{I}^{\dagger}(\varphi)\urcorner$.*

Proof. Let † be

$$p^{\dagger} = \begin{cases} 0 = 1, & \text{if } V_z(p) = 0, \\ \neg\lambda, & \text{if } V_w(p) = 1 \text{ and } V_z(p) = \mathsf{b}, \\ \lambda, & \text{if } V_w(p) = 0 \text{ and } V_z(p) = \mathsf{b}, \\ 0 = 1, & \text{otherwise.} \end{cases}$$

The proof is again by induction on φ both in (i) and (ii). *qed.*

We can now prove Lemma 52.

Proof of Lemma 52. The proofs are again by induction on the positive build up of φ. In cases (i)a, (i)b and (i)c one employs $\circ$. We only provide some details for (i)c mainly because of the peculiar nature of the interaction between $\to$ and $\twoheadrightarrow$. The other claims are easier.

If φ is p, then by assumption either $p^\circ = (0 = 0)$ or $p^\circ = \lambda$, so $\mathsf{DT} \vdash p^\circ$ by the DT axioms and Lemma 41. If φ is $\neg p$, then $p^\circ = (0 = 1)$ or $p^\circ = \lambda$, and in either case $\mathsf{DT} \vdash \neg p^\circ$. If φ is $\Box\psi$ or $\neg\Box\psi$, the claims follow from Lemma 53. If φ is $\psi \twoheadrightarrow \chi$, by Remark 40 it is sufficient to show that $\mathsf{DT} \vdash \mathfrak{I}^\circ(\ulcorner\psi\urcorner) \to \mathfrak{I}^\circ(\ulcorner\chi\urcorner)$. If $\mathcal{M}, w \Vdash_{\mathsf{f3}} \psi \twoheadrightarrow \chi$, then either $\mathcal{M}, w \Vdash_{\mathsf{f3}} \neg\psi$ or $\mathcal{M}, w \Vdash_{\mathsf{f3}} \neg\psi \wedge \chi$. By induction hypothesis, in either case we obtain the desired claim. Similarly, if $\mathcal{M}, z \Vdash_{\mathsf{f3}} \neg(\psi \twoheadrightarrow \chi)$, then $\mathcal{M}, w \Vdash_{\mathsf{f3}} \psi$ and $\mathcal{M}, w \Vdash_{\mathsf{f3}} \neg\chi$, therefore by induction hypothesis $\mathsf{DT} \vdash \mathfrak{I}^\circ(\psi) \wedge \mathfrak{I}^\circ(\neg\chi)$. The claim is then obtained by Remark 40 and by definition of $\mathfrak{I}^\circ$.

For (ii) one employs $\dagger$. The rest is analogous. *qed.*

We finally establish the main result of this section.

PROPOSITION 55.

(i) *Let S be a consistent, first-order extension of* $\mathsf{KF} + \mathsf{CN}$ *in* $\mathcal{L}_{\mathrm{T}}$*, then*

$$\mathsf{M}^{\mathsf{n}} \vdash \varphi \text{ if and only if for all realizations } \star,\ S \vdash \mathfrak{I}^\star(\varphi).$$

(ii) *Let S be a consistent, first-order extension of* WKFC *in* $\mathcal{L}_{\mathrm{T}}$*, then*

$$\mathsf{M}^{\mathsf{w}} \vdash \varphi \text{ if and only if for all realizations } \star,\ S \vdash \mathfrak{I}^\star(\varphi).$$

(iii) *Let S be a consistent, first-order extension of* DT *in* $\mathcal{L}_{\mathrm{T}}$*, then*

$$\mathsf{M}^{\mathsf{f}} \vdash \varphi \text{ if and only if for all realizations } \star,\ S \vdash \mathfrak{I}^\star(\varphi).$$

(iv) *Let S be a consistent, first-order extension of* $\mathsf{KF} + \mathsf{CM}$ *in* $\mathcal{L}_{\mathrm{T}}$*, then*

$$\mathsf{M}^{\mathsf{b}} \vdash \varphi \text{ if and only if for all realizations } \star,\ S \vdash \mathfrak{I}^\star(\varphi).$$

Proof. We consider the case for $\mathsf{KF} + \mathsf{CN}$, the other are analogous. Corollary 50 already gives us the soundness direction. For the completeness direction, if $\mathsf{M}^{\mathsf{n}} \nvdash \varphi$, then by Corollaries 29 and 30 there is a single-rooted mixed, idiosyncratic, consistent faithful model $\mathcal{M}$ such that $\mathcal{M}, w \Vdash_{\mathsf{k3}} \neg\varphi$ for $w \in W$. By Lemma 52 (i)a, $\mathsf{KF} + \mathsf{CN} \vdash \mathfrak{I}^\circ(\neg\varphi)$. Therefore, there is a realization $\circ$ such that $S \nvdash \mathfrak{I}^\circ(\varphi)$ for all $S \supseteq \mathsf{KF} + \mathsf{CN}$. *qed.*

Propositions 55 shows that the modal logics $\mathsf{M}^{\mathsf{n}}, \mathsf{M}^{\mathsf{w}}, \mathsf{M}^{\mathsf{f}}, \mathsf{M}^{\mathsf{b}}$ are maximal in the two senses defined in the introduction: first, there is no consistent modal logic properly extending these logics that is the modal logic of a theory of truth; as a consequence, there is no consistent modal logic properly extending these logics that can be consistently extended with fixed points for all modal formulas in the sense of [Smo85].

5.3. **Weaker base theories.** So far we have considered only theories of truth extending PA. It is natural to ask whether this assumption can be weakened. The short answer is that it can: as long as the base theory supports a satisfactory development of formal syntax, including the structural basis for the diagonal lemma, the arguments above can be employed to determine the modal logics of the Kripke-Feferman theories over the base theory. This is unlike what happens in Solovay's proof of the arithmetical completeness of GL, in which one typically requires exponentiation in the formalization of arbitrary (finite) GL-frames in $\mathrm{I}\Delta_0 + \mathrm{Exp}$.

More precisely, one can consider any $U \supseteq \mathsf{I}\Delta_0 + \Omega_1$, where

$$\Omega_1 := \forall x \exists y\ x^{|x|} = y$$

and $|x| = \lceil \log_2(x+1) \rceil$ stands for the length of the binary representation of x. Since the formal syntax required for the formulation of Kripke-Feferman theories (and, also, Gödel's incompleteness theorems) can be shown to involve only p-time notions and operations [Bus86],[21] and $\mathrm{I}\Delta_0 + \Omega_1$ is enough to prove the totality of all p-time functions [HP98], we can then define the Kripke-Feferman systems $\mathfrak{T}$ on top of such base theories U as prescribed by Definitions 37-39. Models of $\mathfrak{T}[U]$ will then possess a coding scheme [Mos74], and fixed-point semantics can be built over them in the way described in §4.2. Crucially, Liars and truth-tellers will behave in the expected way, so that we can prove analogues of Lemmata 41 and 46. From this point onwards the arguments in sections 5.1 and 5.2 carry over with very little modifications.

What we just said raises the obvious question why we did not present the results above for Kripke-Feferman theories $\mathfrak{T}[U]$, with U an extension of $\mathrm{I}\Delta_0 + \Omega_1$ or S^1_2. One reason is that the status of theories $\mathfrak{T}[U]$ *qua* satisfactory theories of truth is an open question. We know too little concerning their proof-theoretic strength, speed-up properties, and philosophical applications to conceive of them on a par with their standard versions over PA. For instance, Feferman devised KF (and variants thereof) as essentially featuring an open-ended induction schema, which was crucial for characterizing the reflective closure of PA. Restrictions of induction involved in the general definitions of the Kripke-Feferman theories clash with these original projects. However, if one regards of the theories $\mathfrak{T}[U]$ as adequate systems, our arguments can be applied to determine their modal logic.

6. Conclusion

We determined the (propositional) modal logic of Feferman's axiomatizations of Kripke's theory of truth. In the case of systems whose truth-predicates behave according to a paracomplete or paraconsistent three-valued logics, such modal logics amount to the modal logics of all of their consistent extensions, including the sentences satisfied by consistent and complete fixed point models.

In the present paper we did not consider paraconsistent (three- or four-valued) theories of truth based on Weak Kleene or Feferman-Aczel logic. We expect that our results extend to such cases with little modification, but this would need to be verified in detail. In particular, the Witness realization would need to be changed to accommodate the behaviour of the Weak-Kleene disjunction, and the surrounding lemmata would need to be changed accordingly. Similar modifications, although arguably less drastic, are required to establish analogues of the result in §5.2 for the paraconsistent (three-valued) versions of WKFC and DT. It would also be interesting to investigate the modal logics of theories whose truth predicate is sound with respect to supervaluational fixed points.

Perhaps surprisingly the truth-theoretic completeness results of this paper can be lifted to the setting of first-order modal logic, that is, we can determine the first-order modal logics of Kripke-Feferman truth. This contrasts strongly with the situation in the case of provability where it is well known that the first-order logic of provability cannot be axiomatized. The principal reason for this asymmetry is that in the first-order modal logics of truth the quantifiers will commute with the modal operator, which is not the case for quantified provability logic. The details of the truth-theoretic completeness results for first-order modal logic are presented in our companion paper [NS20].

Acknowledgments. We thank Luca Castaldo, Stanislav Speranski for helpful comments on draft version of this paper and Catrin Campbell-Moore for very helpful discussions. We also wish to thank the audiences of the FSB seminar at Bristol,

[21] As better alternative to $\mathsf{I}\Delta_0 + \Omega_1$, one can even employ Buss' S^1_2, which has the advantage of being finitely axiomatized while being equally able to develop syntax adequately.

the AAL Meeting 2018 in Wellington, the 2018 Summer School in Proof Theory in Ghent, Gap10 at Cologne, the Axiomatizing Metatheory Workshop in Salzburg, Logic in London I, the British Logic Colloquium and the Colloquium in Mathematical Philosophy at the MCMP. Johannes Stern's research was supported by the European Research Council (ERC) Starting Grant Truth and Semantics (TRUST, Grant n ∘ 803684).

References

[AB75] Alan R. Anderson and Nuel D. Belnap. *Entailment: The Logic of Relevance and Neccessity, Vol. I.* Princeton University Press, 1975.

[Acz80] Peter Aczel. Frege structures and the notions of proposition, truth and set. In *The Kleene Symposium (Proc. Sympos., Univ. Wisconsin, Madison, Wis., 1978)*, volume 101 of *Stud. Logic Foundations Math.*, pages 31–59. North-Holland, Amsterdam-New York, 1980.

[BdRV01] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, Cambridge, 2001.

[Bur14] John P. Burgess. Friedman and the axiomatization of Kripke's theory of truth. In *Foundational adventures*, volume 22 of *Tributes*, pages 125–148. Coll. Publ., London, 2014.

[Bus86] S. Buss. *Bounded arithmetic*. Bibliopolis, Napoli, 1986.

[Can89] A. Cantini. Notes on formal theories of truth. *Zeitschrift für Logik un Grundlagen der Mathematik*, 35:97–130, 1989.

[CC13] Marcelo Coniglio and Maria Corbalan. Sequent calculi for the classical fragment of bochvar and halldén's nonsense logics. *Electronic Proceedings in Theoretical Computer Science*, 113, 03 2013.

[CD91] James Cain and Zlatan Damnjanovic. On the weak kleene scheme in kripke's theory of truth. *Journal of Symbolic Logic*, 56(4):1452–1468, 1991.

[Che80] Brian F. Chellas. *Modal logic*. Cambridge University Press, Cambridge-New York, 1980. An introduction.

[Cos74] Newton C. A. Da Costa. On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic*, 15(4):497–510, 1974.

[CS20] Luca Castaldo and Johannes Stern. KF, PKF, and Reinhardt's Program, 2020.

[CZ19] Marek Czarnecki and Konrad Zdanowski. A modal logic of a truth definition for finite models. *Fund. Inform.*, 164(4):299–325, 2019.

[Fef84] S. Feferman. Towards useful type-free theories i. *Journal of Symbolic Logic*, 49(1):75–111, 1984.

[Fef91] S. Feferman. Reflecting on incompleteness. *Journal of Symbolic Logic*, 56: 1–49, 1991.

[Fef08] Solomon Feferman. Axioms for determinateness and truth. *Rev. Symb. Log.*, 1(2):204–217, 2008.

[FHKS15] M. Fischer, V. Halbach, J. Kriener, and J. Stern. Axiomatizing semantic theories of truth? *The Review of Symbolic Logic*, 8(2):257–278, 2015.

[FS87] Harvey Friedman and Michael Sheard. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33(1):1–21, 1987.

[Fuj10] Kentaro Fujimoto. Relative truth definability of axiomatic truth theories. *Bulletin of Symbolic Logic*, 16(3):305–344, 2010.

[GB93] A. Gupta and N. Belnap. *The Revision Theory of Truth*. MIT Press, 1993.

[Hal94] Volker Halbach. A system of complete and consistent truth. *Notre Dame J. Formal Logic*, 35(3):311–327, 1994.

[Hal14] V. Halbach. *Axiomatic theories of truth. Revised edition.* Cambridge University Press, 2014.

[HH06] V. Halbach and L. Horsten. Axiomatizing Kripke's theory of truth in partial logic. *Journal of Symbolic Logic*, 71: 677–712, 2006.

[HP98] Petr Hájek and Pavel Pudlák. *Metamathematics of first-order arithmetic*. Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1998. Second printing.

[JT96] Jan Jaspars and Elias Thijsse. Fundamentals of partial modal logic. In Partick Doherty, editor, *Partiality, Modality, Nonmonotonicity*, pages 111–141. CSLI, 1996.

[Kle52] Stephen Cole Kleene. *Introduction to Metamathematics*. North Holland, 1952.

[Kri75] S. Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–712, 1975.

[Loe55] M. H. Loeb. Solution of a problem of Leon Henkin. *J. Symbolic Logic*, 20:115–118, 1955.

[Mar84] Robert L. Martin. On representing true-in-l'in l robert l. martin and peter w. woodruff. In Robert L. Martin, editor, *Recent Essays on Truth and the Liar Paradox*, page 47. Oxford University Press, 1984.

[Mos74] Yiannis N. Moschovakis. *Elementary induction on abstract structures*. North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York, 1974. Studies in Logic and the Foundations of Mathematics, Vol. 77.

[Nic18] Carlo Nicolai. Provably true sentences across axiomatizations of kripke's theory of truth. *Studia Logica*, 106(1):101–130, 2018.

[NS20] C. Nicolai and J. Stern. First-order modal logics of truth. *Unpublished Manuscript*, 2020.

[OS16] S.P. Odintsov and Speranski S.O. The lattice of Belnapian modal logics: special extensions and counterparts. *Logic and Logical Philosophy*, 25(1):3–33, 2016.

[OS20] S.P. Odintsov and Speranski S.O. Belnap–Dunn modal logics: Truth constants vs. truth values. *The Review of Symbolic Logic*, 13(2):416–435, 2020.

[OW10] Sergei P Odintsov and Heinrich Wansing. Modal logics with Belnapian truth values. *Journal of Applied Non-Classical Logics*, 20(3):279–301, 2010.

[Pak17] Fedor Pakhomov. Solovay's completeness without fixed points. In *Logic, language, information, and computation*, volume 10388 of *Lecture Notes in Comput. Sci.*, pages 281–294. Springer, Berlin, 2017.

[Pri08] Graham Priest. *An Introduction to Non-Classical Logic: From If to Is.* Cambridge University Press, 2008.

[Seg71] Krister Segerberg. *An essay in classical modal logic. Vols. 1, 2, 3.* Filosofiska Föreningen och Filosofiska Institutionen vid Uppsala Universitet, Uppsala, 1971. Filosofiska Studier, No. 13.

[Smo85] C. Smoryński. *Self-reference and modal logic*. Universitext. Springer-Verlag, New York, 1985.

[Sol76] Robert M. Solovay. Provability interpretations of modal logic. *Israel J. Math.*, 25(3-4):287–304, 1976.

[Spe17] Stanislav O. Speranski. Notes on the computational aspects of kripke's theory of truth. *Studia Logica*, 105(2):407–429, 2017.

[Sta15] Shawn Standefer. Solovay-type theorems for circular definitions. *Rev. Symb. Log.*, 8(3):467–487, 2015.

[Ste15] Johannes Stern. *Toward predicate approaches to modality*, volume 44. Springer, 2015.

[Tar55] Alfred Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific J. Math.*, 5:285–309, 1955.

[Tar56] A. Tarski. Der Wahrhetisbegriff in den formalisierten Sprachen. In *Logic, semantics, metamathematics*, pages 152–278. Clarendon Press, Oxford, 1956.

[Vis89] Albert Visser. Semantics and the liar paradox. *Handbook of Philosophical Logic*, 4(1):617–706, 1989.

Appendix A

The basic nonclassical system we are interested in is the four-valued nonclassical logic known as first-degree entailment [AB75].

Definition 56 (FDE).

$$(\textsc{ref})\quad \Gamma, \varphi \Rightarrow \varphi, \Delta \text{ for } \varphi \text{ a literal} \qquad (\textsc{cut})\ \frac{\Gamma \Rightarrow \Delta, \varphi \qquad \varphi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}$$

$$(\bot)\quad \Gamma, \bot \Rightarrow \Delta \qquad (\top)\quad \Gamma \Rightarrow \top, \Delta$$

$$(\textsc{dn-l})\ \frac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma, \neg\neg\varphi \Rightarrow \Delta} \qquad (\textsc{dn-r})\ \frac{\Gamma \Rightarrow \varphi, \Delta}{\Gamma \Rightarrow \neg\neg\varphi, \Delta}$$

$$(\neg\wedge\textsc{l})\ \frac{\Gamma, \neg\varphi \Rightarrow \Delta \qquad \Gamma \Rightarrow \neg\psi, \Delta}{\Gamma, \neg(\varphi \wedge \psi) \Rightarrow \Delta} \qquad (\neg\wedge\textsc{r})\ \frac{\Gamma \Rightarrow \neg\varphi_i, \Delta}{\Gamma \Rightarrow \Delta, \neg(\varphi_0 \wedge \varphi_1)}\ i = 0, 1$$

$$(\wedge\textsc{l})\ \frac{\Gamma, \varphi_i \Rightarrow \Delta}{\Gamma, \varphi_0 \wedge \varphi_1 \Rightarrow \Delta} \qquad (\wedge\textsc{r})\ \frac{\Gamma \Rightarrow \varphi, \Delta \qquad \Gamma \Rightarrow \psi, \Delta}{\Gamma \Rightarrow \Delta, \varphi \wedge \psi}$$

$$(\neg\vee\mathrm{L})\ \frac{\Gamma, \neg\varphi_i \Rightarrow \Delta}{\Gamma, \neg(\varphi_0 \vee \varphi_1) \Rightarrow \Delta} \qquad (\neg\vee\mathrm{R})\ \frac{\Gamma \Rightarrow \neg\varphi, \Delta \qquad \Gamma \Rightarrow \neg\psi, \Delta}{\Gamma \Rightarrow \neg(\varphi \vee \psi), \Delta}$$

$$(\vee\mathrm{L})\ \frac{\Gamma, \varphi \Rightarrow \Delta \qquad \Gamma, \psi \Rightarrow \Delta}{\Gamma, \varphi \vee \psi \Rightarrow \Delta} \qquad (\vee\mathrm{R})\ \frac{\Gamma \Rightarrow \varphi_i, \Delta}{\Gamma \Rightarrow \varphi_0 \vee \varphi_1, \Delta}$$

FDE can be considered to be the basis of well-known three-valued paracomplete or paraconsistent systems [Kle52, Cos74].

Definition 57.

(i) KS3 *is obtained by adding to* FDE *the sequent:*

$$(\mathrm{SYM}) \qquad \Gamma, \varphi, \neg\varphi \Rightarrow \psi, \neg\psi, \Delta$$

(ii) K3 *is obtained by adding to* FDE *the rule:*

$$(\neg\mathrm{L})\ \frac{\Gamma \Rightarrow \Delta, \varphi}{\neg\varphi, \Gamma \Rightarrow \Delta}$$

(iii) LP *is obtained by adding to* FDE *the rule:*

$$(\neg\mathrm{R})\ \frac{\varphi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg\varphi}$$

The next logic we consider Weak Kleene logic. Our axiomatization is a variant of the one that can be found in [CC13].

Definition 58 (Weak Kleene, B3).

$$(\mathrm{REF}) \quad \Gamma, p \Rightarrow p, \Delta \qquad (\mathrm{CUT})\ \frac{\Gamma \Rightarrow \Delta, \varphi \qquad \varphi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}$$

$$(\bot) \quad \Gamma, \bot \Rightarrow \Delta \qquad (\top) \quad \Gamma \Rightarrow \top, \Delta$$

$$(\neg\mathrm{L})\ \frac{\Gamma \Rightarrow \varphi, \Delta}{\Gamma, \neg\varphi \Rightarrow \Delta} \qquad (\neg\mathrm{R})\ \frac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \neg\varphi, \Delta}\ \textit{with } \mathrm{Prop}(\varphi) \subseteq \mathrm{Prop}(\Gamma)$$

$$(\wedge\mathrm{L})\ \frac{\Gamma, \varphi_i \Rightarrow \Delta}{\Gamma, \varphi_0 \wedge \varphi_1 \Rightarrow \Delta} \qquad (\wedge\mathrm{R})\ \frac{\Gamma \Rightarrow \varphi, \Delta \qquad \Gamma \Rightarrow \psi, \Delta}{\Gamma \Rightarrow \varphi \wedge \psi, \Delta}$$

$$(\vee\mathrm{L})\ \frac{\Gamma, \varphi \Rightarrow \Delta \qquad \Gamma, \psi \Rightarrow \Delta}{\Gamma, \varphi \vee \psi \Rightarrow \Delta} \qquad (\vee\mathrm{R})\ \frac{\Gamma \Rightarrow \varphi_i, \Delta}{\Gamma \Rightarrow \varphi_0 \vee \varphi_1, \Delta}\ \textit{with } \mathrm{Prop}(\varphi_0, \varphi_1) \subseteq \mathrm{Prop}(\Gamma)$$

Remark 59.

(i) In FDE (and extensions thereof), (REF) is defined for literals, whereas in B3 it is defined for propositional variables only. This is because the negation rules of B3 enable us to straightforwardly derive reflexivity for literals.

(ii) The restriction on the rules ($\wedge$R) and ($\neg$R) can be seen as 'forcing' a determinate truth value on the principal formulas.

The last logic we consider is the extension of Weak Kleene considered (semantically) by Peter Aczel for his Frege Structures [Acz80], and Feferman in [Fef08].

Definition 60 (Feferman Logic, F3). *The language* $\mathcal{L}_0^{\twoheadrightarrow}$ *of* F3*, besides the connectives of* B3*, features a special conditional* $\twoheadrightarrow$*. The rules of* F3 *are the the rules of* B3 *plus:*

$$(\twoheadrightarrow\text{L})\ \frac{\Gamma \Rightarrow \Delta, \varphi \qquad \psi, \Gamma \Rightarrow \Delta}{\Gamma, \varphi \twoheadrightarrow \psi \Rightarrow \Delta}$$

$$(\twoheadrightarrow\text{R1})\ \frac{\Gamma \Rightarrow \neg\varphi, \Delta}{\Gamma \Rightarrow \varphi \twoheadrightarrow \psi, \Delta} \qquad (\twoheadrightarrow\text{R2})\ \frac{\Gamma, \varphi \Rightarrow \psi, \Delta}{\Gamma \Rightarrow \varphi \twoheadrightarrow \psi, \Delta}$$

$$\textit{with } \mathrm{Prop}(\varphi, \psi) \subseteq \mathrm{Prop}(\Gamma)$$

Remark 61. ($\twoheadrightarrow$R2) is derivable in B3 for the material conditional defined by $\neg$ and $\vee$.

Appendix B: Completeness of the Modal Nonclassical Systems

As usual, the soundness direction is straightforward. The main idea of the completeness proof is to modify the standard Henkin strategy by replacing the notion of maximally consistent set with the one of *saturated set.*

Definition 62 ($S_{\blacksquare}$-Saturated set [JT96]). *For* $S_{\blacksquare}$ *as above, a set* Γ *of* $\mathcal{L}_{\Box}$*-sentences is* $S_{\blacksquare}$*-saturated iff for all finite* $\Delta \subseteq \mathrm{Sent}_{\mathcal{L}_{\Box}}$*: if* $S_{\blacksquare} \vdash \Gamma \Rightarrow \Delta$*, then* $\Gamma \cap \Delta \neq \varnothing$*.*

Lemma 63. *If* $S_{\blacksquare} \nvdash \Gamma \Rightarrow \Delta$*, then there is a* $S_{\blacksquare}$*-saturated* $\Gamma^* \supseteq \Gamma$ *such that* $\Gamma^* \cap \Delta = \varnothing$*.*

Proof sketch. Starting with an enumeration of $\mathcal{L}_{\Box}$-sentences in which every sentence occurs infinitely many times, one defines:

$$\Gamma_0 =: \Gamma$$

$$\Gamma_{n+1} =: \begin{cases} \Gamma_n \cup \{\varphi_n\}, & \text{if } S_{\blacksquare} \vdash \Gamma_n, \varphi_n \Rightarrow \Theta \text{ entails } \Theta \cap (\mathrm{Sent}_{\mathcal{L}_{\Box}} \setminus \Delta) \neq \varnothing \\ & \text{ for all finite } \Theta \subseteq \mathrm{Sent}_{\mathcal{L}_{\Box}}, \\ \Gamma_n & \text{otherwise;} \end{cases}$$

$$\Gamma^* := \bigcup_{n \in \omega} \Gamma_n$$

Now $\Gamma^* \subseteq \mathrm{Sent}_{\mathcal{L}_{\Box}} \setminus \Delta$. Therefore, $\Gamma^* \cap \Delta = \varnothing$. It remains to be shown that Γ^* is $S_{\blacksquare}$-saturated. If $S_{\blacksquare} \vdash \Gamma^* \Rightarrow \Theta$ for some Θ then, since deductions are finite, there is an n and a finite $\Theta_0 \subseteq \Theta$ such that $S_{\blacksquare} \vdash \Gamma_n \Rightarrow \Theta_0$. By induction on the size of the finite set $\Theta_0 \cap (\mathrm{Sent}_{\mathcal{L}_{\Box}} \setminus \Delta)$ – [JT96, Lemma 4.3] –, one obtains that $\Gamma^* \cap \Theta \neq \varnothing$. *qed.*

Canonical models are then constructed from saturated sets is the usual way. However, in contrast to the classical case it no longer suffices to define the accessibility relation $z_0 R z_1$ simply by requiring z_1 to be a superset of $\{\varphi \mid \Box\varphi \in z_0\}$. Rather we also need to stipulate that $z_1 \subseteq \{\varphi \mid \Diamond\varphi \in z_0\}$.[22] In the classical setting these two conditions are equivalent since our worlds are assumed to be maximally consistent.

Definition 64 (Canonical frame). *For* S *as above, the canonical frame for* $S_{\blacksquare}$ (Z_S, R_S) *is specified by:*

$$Z^S := \{z \mid z \textit{ is } S_{\blacksquare}\textit{-saturated}\}$$

$$R^S z_0 z_1 :\Leftrightarrow \{\varphi \mid \Box\varphi \in z_0\} \subseteq z_1 \subseteq \{\varphi \mid \Diamond\varphi \in z_0\}$$

[22] Recall that in our nonclassical context $\Diamond := \neg\Box\neg$, whereas this will not be true in the context of our classical modal logic.

The canonical model is obtained from the canonical frame by extending it with a suitable evaluation. The details of the evaluations vary depending on the kind of saturated set we are considering. We let, for $S \in \{\mathsf{FDE}, \mathsf{K3}, \mathsf{LP}, \mathsf{B3}, \mathsf{F3}, \mathsf{KS3}\}$:

$$V_z^S(p) = \begin{cases} 1 & \text{if } p \in z \text{ and } \neg p \notin z \\ 0 & \text{if } \neg p \in z \text{ and } p \notin z \\ \mathsf{b} & \text{if } p \in z \text{ and } \neg p \in z \\ \mathsf{n} & \text{otherwise} \end{cases}$$

Definition 65 (Canonical model). *For $S \in \{\mathsf{K3}, \mathsf{B3}, \mathsf{F3}, \mathsf{LP}, \mathsf{FDE}, \mathsf{KS3}\}$, the canonical model $\mathcal{M}^S$ for $S_\blacksquare$ is the triple (Z^S, R^S, V^S).*

Lemma 66 (Existence). *Let z_0 and z_1 be $S_\blacksquare$-saturated. Then the following implications hold:*

(i) *if $\{\varphi \mid \Box\varphi \in z_0\} \subseteq z_1$, then there is an $S_\blacksquare$-saturated $z \subseteq z_1$ such that $R^S z_0 z$;*

(ii) *if $z_1 \subseteq \{\varphi \mid \Diamond\varphi \in z_0\}$, then there is an $S_\blacksquare$-saturated $z \supseteq z_1$ such that $R^S z_0 z$.*

Proof. We start with (i). Since obviously

$$\mathrm{Sent}_{\mathcal{L}_\Box} \setminus (\mathrm{Sent}_{\mathcal{L}_\Box} \setminus (z_1 \cap \{\varphi \mid \Diamond\varphi \in z_0\})) = z_1 \cap \{\varphi \mid \Diamond\varphi \in z_0\}))$$

one starts by noticing that

$$S_\blacksquare \nvdash \{\varphi \mid \Box\varphi \in z_0\} \Rightarrow \mathrm{Sent}_{\mathcal{L}_\Box} \setminus (z_1 \cap \{\varphi \mid \Diamond\varphi \in z_0\}).$$

This is because, if

$$S_\blacksquare \vdash \{\varphi \mid \Box\varphi \in z_0\} \Rightarrow \mathrm{Sent}_{\mathcal{L}_\Box} \setminus (z_1 \cap \{\varphi \mid \Diamond\varphi \in z_0\}),$$

then

$$S_\blacksquare \vdash \{\varphi \mid \Box\varphi \in z_0\} \Rightarrow \Theta,$$

for some finite $\Theta \subseteq \mathrm{Sent}_{\mathcal{L}_\Box} \setminus (z_1 \cap \{\varphi \mid \Diamond\varphi \in z_0\})$. Since $\{\varphi \mid \Box\varphi \in z_0\} \subseteq z_1$ and z_1 is $S_\blacksquare$-saturated, $z_1 \cap \Theta \neq \varnothing$. So we can divide up Θ in such a way that

$$S_\blacksquare \vdash \{\varphi \mid \Box\varphi \in z_0\} \Rightarrow \bigvee(\Theta \setminus z_1), \Theta \cap z_1.$$

By the $S_\blacksquare$ rules,

$$S_\blacksquare \vdash z_0 \Rightarrow \Box \bigvee(\Theta \setminus z_1), \Diamond(\Theta \cap z_1).$$

Since z_0 is $S_\blacksquare$-saturated, either $\Box \bigvee(\Theta \setminus z_1) \in z_0$, or $\Diamond(\Theta \cap z_1) \cap z_0 \neq \varnothing$. If the former, then $\bigvee(\Theta \setminus z_1) \in z_1$, which is impossible. If the latter, then $z_1 \cap \{\varphi \mid \Diamond\varphi \in z_0\} \cap \Theta \neq \varnothing$, which is also impossible.

Therefore, by Lemma 63, we can construct an $S_\blacksquare$-saturated z such that

$$\{\varphi \mid \Box\varphi \in z_0\} \subseteq z \subseteq z_1 \cap \{\varphi \mid \Diamond\varphi \in z_0\},$$

which yields of course the claim.

The proof of (ii) is similar to the previous case. Since z_0 is $S_\blacksquare$-saturated,

$$S_\blacksquare \nvdash \{\varphi \mid \Box\varphi \in z_0\}, z_1 \Rightarrow \mathrm{Sent}_{\mathcal{L}_\Box} \setminus \{\varphi \mid \Diamond\varphi \in z_0\}.$$

Again by Lemma 63 there is a z such that

$$\{\varphi \mid \Box\varphi \in z_0\} \cup z_1 \subseteq z \subseteq \{\varphi \mid \Diamond\varphi \in z_0\}$$

as desired. *qed.*

Lemma 67 (Truth Lemma). *Let $z \in Z^S$ for $S \in \{\mathsf{K3}, \mathsf{B3}, \mathsf{LP}, \mathsf{FDE}, \mathsf{KS3}\}$. Then for all $\varphi \in \mathcal{L}_\Box$:*

$$\mathcal{M}^S, z \Vdash_s \varphi \text{ if and only if } \varphi \in z.$$

Proof Sketch. By induction on the positive complexity of φ. There are two non-trivial cases. The first is when φ is of the form $\Box\psi$. The right-to-left direction is obtained by induction hypothesis. In the left-to-right direction, starting with $\Box\psi \notin z$, one finds a $S_{\blacksquare}$-saturated set $z_1 \supseteq \{\varphi \mid \Box\varphi \in z\}$. By the first part of Lemma 66, there is an $S_{\blacksquare}$-saturated $z_0 \subseteq z_1$ such that $R^S zz_0$. Since $\varphi \notin z_0$, by induction hypothesis one obtains that $\mathcal{M}^S, z_0 \nVdash_s \varphi$, as required.

The second non-trivial case, when φ is $\neg\Box\psi$, is symmetric to the previous one and employs the second part of Lemma 66. For the right-to-left direction, suppose $\neg\Box\psi \in v$. One then notices that

$$S_{\blacksquare} \nvdash \{\varphi \mid \Box\varphi \in v\}, \neg\psi \Rightarrow \varnothing \tag{4}$$

Therefore, by Lemma 63, we can find a saturated $z_0 \supseteq \{\varphi \mid \Box\varphi \in v\}, \neg\psi$. By Lemma 66, there is also a $z_1 \supseteq z_0$ with $R^S zz_1$. So, $\neg\psi \in z_1$. The claim then follows by induction hypothesis. *qed.*

We can finally prove the adequacy of our systems.

Proof of Prop. 8. The soundness direction is obtained by a straightforward induction on the length of the proof in $S_{\blacksquare}$.

For the completeness direction, one assumes that $S_{\blacksquare} \nvdash \Gamma \Rightarrow \Delta$ and finds, by Lemma 63 an $S_{\blacksquare}$-saturated $z \supseteq \Gamma$ such that $\Delta \cap z = \varnothing$. By the truth Lemma, $\mathcal{M}^S, z \Vdash_s \gamma$ for all $\gamma \in \Gamma$ and $\mathcal{M}^S, z \nVdash_s \delta$ for any $\delta \in \Delta$, as required. *qed.*

King's College London and University of Bristol
Email address: `carlo.nicolai@kcl.ac.uk; johannes.stern@bristol.ac.uk`